

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO



DISEÑO DE UNA METODOLOGÍA MULTICRITERIO
DE APOYO A LA DECISIÓN PARA LA GESTIÓN DE
LA PERMANENCIA ESTUDIANTIL DE EDUCACIÓN
SUPERIOR

POR

LUISA FERNANDA ARRIETA MATOS

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE

DOCTORADO EN INGENIERÍA

CON ESPECIALIDAD EN INGENIERÍA DE SISTEMAS

JULIO 2021

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO



DISEÑO DE UNA METODOLOGÍA MULTICRITERIO
DE APOYO A LA DECISIÓN PARA LA GESTIÓN DE
LA PERMANENCIA ESTUDIANTIL DE EDUCACIÓN
SUPERIOR

POR

LUISA FERNANDA ARRIETA MATOS

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE

DOCTORADO EN INGENIERÍA

CON ESPECIALIDAD EN INGENIERÍA DE SISTEMAS

JULIO 2021



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica
Subdirección de Estudios de Posgrado

Los miembros del Comité de Tesis recomendamos que la Tesis «Diseño de una metodología multicriterio de apoyo a la decisión para la gestión de la permanencia estudiantil de educación superior», realizada por la alumna LUISA FERNANDA ARRIETA MATOS, con número de matrícula 001937492, sea aceptada para su defensa como requisito para obtener el grado de Doctorado en Ingeniería con Especialidad en Ingeniería de Sistemas.

El Comité de Tesis

Dr. Fernando López Irragarri
Director

Dr. Paul Sanmartín Mendoza
Co-Director

Dr. Romeo Sánchez Nigenda
Revisor

Dr. Marco González Viquez
Revisor

Dra. Iris Abril Martínez Salazar
Revisor

Dra. Yasmín Águeda Ríos Solís
Revisor

Dr. Simón Martínez Martínez
Subdirector de Estudios de Posgrado

San Nicolás de los Garza, Nuevo León, Julio 2021



Ciudad Universitaria Pedro de Alba s/n, C.P. 66455. A.P. 076 Suc. "F"
San Nicolás de los Garza, Nuevo León, México. Tels: (81) 8332 0903 /
Conn.: 8329 4020 / Fax: (81) 8332 0904

A mis hijos

Para ustedes y por ustedes.

Los amo.

AGRADECIMIENTOS

Agradezco a Dios y todos aquellos quienes aportaron un granito de arena en la realización de esta tesis.

Mi familia y Director. Gracias.

ÍNDICE GENERAL

Agradecimientos	v
1. Introducción	1
1.1. CONTEXTO	2
1.2. ANTECEDENTES	3
1.3. DEFINICIÓN DEL PROBLEMA CIENTÍFICO	4
1.4. HIPOTESIS	5
1.5. OBJETIVOS	6
1.6. OBJETIVOS ESPECÍFICOS	6
1.7. JUSTIFICACIÓN	6
1.8. NOVEDAD CIENTÍFICA	7
1.9. ESTRUCTURA DEL DOCUMENTO	7
2. MARCO TEÓRICO Y ESTADO DEL ARTE	8
2.1. DESERCIÓN ESTUDIANTIL	8
2.1.1. ABANDONO ESCOLAR / DESERCIÓN	8

2.2. CIENCIA DE DATOS	9
2.2.1. METODOLOGÍA FUNDAMENTAL PARA LA CIENCIA DE DATOS IBM	11
2.2.2. CRISP - DM	14
2.2.3. APRENDIZAJE AUTOMATIZADO	15
2.3. TEORIA DE APOYO A LA DECISIÓN	19
2.3.1. EL MODELO DE RACIONALIDAD LIMITADA	19
2.3.2. CONJUNTOS APROXIMADOS. ROUGHT SET	25
2.4. REVISIÓN DEL ESTADO DEL ARTE	30
2.4.1. DESERCIÓN O PERMANENCIA ESTUDIANTIL	30
2.4.2. ANÁLISIS DE DATOS EDUCATIVOS	32
3. DESCRIPCIÓN DETALLADA DEL PROBLEMA	37
3.1. PLANTEAMIENTO DEL PROBLEMA	37
3.2. CARACTERÍSTICAS DEL PROBLEMA	39
4. METODOLOGÍA PROPUESTA	42
4.1. MODELO PROPUESTO PARA GESTIÓN DE LA PERMANEN- CIA DE CALIDAD EN UNA INSTITUCIÓN DE EDUCACIÓN SU- PERIOR	42
4.2. ENFOQUE METODOLÓGICO	50
4.2.1. REQUERIMIENTOS PARA LA INTERVENCIÓN	51
4.3. ETAPA DE DIAGNÓSTICO	52

4.3.1. DIAGNÓSTICO COMO PROBLEMA DE DECISIÓN MULTICRITERIO	52
4.4. ETAPA DE INTERVENCIÓN	62
4.4.1. DEFINICIÓN Y ESTRUCTURA DEL PROBLEMA DE DECISIÓN. PASO 1	63
4.4.2. CONSTRUCCIÓN DE LA INSTANCIA DE DATOS DEL PROBLEMA. PASO 2	69
4.4.3. FORMULACIÓN DE LA RECOMENDACIÓN. (PASO 3) . .	74
5. APLICACIÓN DE LA METODOLOGÍA A UN CASO DE ESTUDIO	80
5.1. DESCRIPCIÓN GENERAL DE LA UNIVERSIDAD SIMÓN BOLÍVAR USB	80
5.2. FASE DE DIAGNÓSTICO	86
5.2.1. APLICACIÓN DE LA FASE DE DIAGNÓSTICO A LA USB	86
5.2.2. FORMULACIÓN DE LA RECOMENDACIÓN DE LA FASE DE DIAGNÓSTICO	94
5.3. FASE DE INTERVENCIÓN	97
5.3.1. DEFINICIÓN Y ESTRUCTURA DEL PROBLEMA DE DECISIÓN(PASO 1)	97
5.3.2. CONSTRUCCIÓN DE LA INSTANCIA DE DATOS DEL PROBLEMA (PASO 2)	101
5.3.3. ANÁLISIS EXPLORATORIO DE DATOS	107

5.3.4. CONSTRUCCIÓN DE LAS ESCALAS DE PREFERENCIA DE LOS ATRIBUTOS.	111
5.3.5. FORMULACIÓN DE LA RECOMENDACIÓN.(PASO 3) . .	119
6. CONCLUSIONES	120
6.0.1. CONCLUSIONES	120
6.0.2. TRABAJO FUTURO	121
A. ANEXO 1	123
A.1. TABLA DE EVALUACIÓN	123
B. ANEXO 2	125
B.1. ORIGEN DE DATOS CASO DE ESTUDIO	125
C. ANEXO 3	128
C.1. RESULTADOS DE ESTADISTICAS EN R	128
D. ANEXO 4	139
D.1. ESTADISTICAS DE RAPID MINER	139

ÍNDICE DE FIGURAS

2.1. Científico de Datos y su entorno. Moreno-Salinas	10
2.2. Metodología Fundamental para la ciencia de datos(IBM)Rollins . . .	11
2.3. Resumen Metodología. Propio	15
2.4. Descripción Metodología Simon	20
3.1. Poblacion de graduados OCDE 2015 2018.	38
3.2. Factores Incidentes. Ministerio de Educación Nacional Colombiano . .	40
4.1. Modelo propuesto por el Ministerio de Educación Nacional Colom- biano 2012	43
4.2. Eje de Sensibilización	46
4.3. Eje Cultura de Información.	46
4.4. Eje Mejoramiento Calidad Académica P1	47
4.5. Eje Mejoramiento Calidad Académica P2	47
4.6. Eje Trabajo en grupo con Instituciones de Educación Media	48
4.7. Eje Programa de apoyo a estudiantes	48
4.8. Eje Compromiso con el núcleo familiar	49

4.9. Eje Gestión de Recursos	49
4.10. Eje Trabajo Colaborativo con Instituciones de educación superior . .	49
4.11. Jerarquía de Criterios	55
5.1. Paso del tiempo Universidad Simón Bolívar. Colombia	81
5.2. Matriculas Pregrado	82
5.3. Ranking Universidad Simón Bolívar	83
5.4. Universidad Simón Bolívar en Cifras. 2020	84
5.5. Porcentaje de estudiantes en las Categorías de Deserción	85
5.6. Fase Diagnóstico	86
5.7. Organigrama	98
5.8. Factores a considerar para la deserción	99
5.9. Comparativo Resultados Modelos	115
5.10. Principales razones de deserción	117
C.1. Comparativo Desertor no desertor y ayudas	128
C.2. Comparativo Calificaciones: Excelente, Bueno, Aprobado, Reprobado	129
C.3. Comparativo Categorías de deserción: Académico, Institucional, Económico, Individual(Antropológico)	129
C.4. Comparativo Cohortes Desertores Programas	130
C.5. Comparativo Convenios	131
C.6. Comparativo Estados	131

C.7. Comparativo Estrategia	132
C.8. Comparativo Estrato	132
C.9. Comparativo Género	133
C.10.Comparativo Desertores por periodo	134
C.11.Comparativo Desertores por periodo de ingreso	134
C.12.Comparativo Programas y desertores	135
C.13.Comparativo vulnerabilidad	135
C.14.Comparativo Desertores por año	136
C.15.Comparativo Año de ingreso vs Desertor	136
C.16.Comparativo Años desde la última matrícula	137
C.17.Comparativo Graduados y desertores	138
C.18.Comparativo Cohortes Desertores Programas	138
D.1. Apoyo	140
D.2. Calificaciones	140
D.3. Convenios	141
D.4. Estados	141
D.5. Estrategias	142
D.6. Estratos	142
D.7. Facultades	143
D.8. Isdesertor	144

D.9. Períodos	144
D.10.Vulnerable	145
D.11.YEAR	146
D.12.Year.Desde.Última.Matrícula	146
D.13.Calificaciones	147
D.14.DesertorSINO	147
D.15.Estado	148
D.16.Estrategias	148
D.17.Facultad	149
D.18.Género	149
D.19.Período	150
D.20.Período de Ingreso	150
D.21.Vulnerable	151
D.22.Year	151
D.23.Years.última matrícula	152

ÍNDICE DE TABLAS

1.1. Tabla con Porcentajes de deserción en Colombia	4
2.1. Tabla Atributos del automóvil	26
3.1. Índice de abandono de estudiantes en educación terciaria a 2017. . . .	39
5.1. Matriculados con atención.	104
5.2. Matriculados sin atención.	105
5.3. Desertores con atención	105
5.4. Desertores sin atención.	106
5.5. Porcentajes de Categorías incidentes en la deserción	107
5.6. Distribución de porcentaje de estudiantes por estratos	107
5.7. Porcentaje de estudiantes por Categorías de deserción y Estratos . . .	108
5.8. Tasa de deserción por programas académicos.	109
5.9. Fichero consolidado	111
5.10. Comparativo resultados de Modelos	116
5.11. Selección de features por mayoría Simple	118

5.12. Selección de features por mayoría Simple	119
--	-----

CAPÍTULO 1

INTRODUCCIÓN

En este capítulo se presenta el diseño de la investigación de este trabajo de tesis. El problema científico que se investiga es la relación de factores que inciden en la deserción (académicos, individuales o antropológicos, institucionales, económicos), en las instituciones de educación superior para la definición de una metodología de gestión a la permanencia, donde analizan los datos como un conjunto y permita tomar decisiones frente a la implementación de estrategias de acompañamiento efectivas y eficaces.

A continuación, se presenta la estructura de este capítulo. En la Sección 1.1 se analiza brevemente el fenómeno de la deserción estudiantil en las instituciones de educación superior en Colombia y el mundo. En la Sección 1.2 se describe de manera general el problema científico que se aborda en esta investigación, se discuten antecedentes relevantes del problema. En la sección 1.3 se presenta la justificación del problema de investigación, en la Sección 1.4 se presenta la hipótesis de la investigación. En las Secciones 1.5 y 1.6 se presentan el objetivo general y los objetivos específicos que rigen esta investigación. Luego, se presenta la novedad científica en la Sección 1.7. En la Sección 1.8 se describe el contenido de cada uno de los capítulos que conforman este documento.

1.1 CONTEXTO

La deserción estudiantil es un fenómeno estudiado hace varios años, desde los diferentes niveles de educación ofrecidos en el mundo, en especial porque incide de manera directa en el desarrollo de un país. En Colombia, de acuerdo con Brunner (2016), la educación es probablemente la vía mas efectiva para reducir la inequidad y alcanzar igualdad de condiciones para la generación de ingresos, ya que a través de una educación equitativa, de calidad y con pertinencia la población puede acceder con mayor facilidad a un empleo formal o emprender con más posibilidades de éxito una iniciativa empresarial. Lo anterior, teniendo también presente que la educación es la herramienta más poderosa para promover la movilidad social y para la construcción de equidad de acuerdo al Plan Nacional de desarrollo colombiano 2018 - 2022 (PND (2018)), dando lugar a que los logros alcanzados en las metas de desarrollo económico y social están directamente relacionados con la inversión que se realiza en la educación, entre otros, y por ello se hacen esfuerzos para mejorar la cobertura y calidad de un sistema educativo que permita el acceso de una mayor población a la formación superior. Sin embargo, el cumplimiento de las metas, se ve afectado cuando no logra culminarse el proyecto formativo de los universitarios. Sus consecuencias permean también lo personal y lo familiar en el ámbito del bienestar psicológico y económico. Es por ello que uno de los objetivos del gobierno nacional colombiano, es brindar educación inicial con enfoque de atención integral, en el marco de la Ley 1804 de 2016 de la República de Colombia (2016) ya que considera, que pese a todos los avances que ha logrado el país en los últimos años para mejorar el acceso con calidad a la educación, aún persisten rezagos en cobertura, permanencia y calidad en todos los niveles. Así mismo continua en el proceso de apoyo para el fomento a la permanencia y graduación oportuna de los estudiantes, apostando a la mitigación del fenómeno de abandono estudiantil, que alienta a las Instituciones de Educación Superior a reducir la deserción. Sin embargo, dada su complejidad, las acciones propuestas no logran la reducción deseada del fenómeno ya que en el aban-

dono convergen un sinnúmero de factores o determinantes no sujetos al control por parte de las instituciones educativas, de acuerdo a estudios realizados en el marco del Proyecto alfa-guía Issuu (16 de Diciembre de 2014) desde el año 2013 hasta la fecha. En el presente trabajo, se propone una metodología multicriterio de apoyo a la decisión para la gestión de Permanencia y éxito estudiantil, que ayuda a definir acciones a seguir para disminuir el abandono de los estudiantes, durante su estancia en la universidad. El desarrollo de la investigación se subdivide en los apartados que corresponden al método científico, considerando un enfoque cuantitativo de análisis. Se proyecta además que la metodología, sea utilizada como referente para las Instituciones de Educación Superior que pertenezcan al grupo de referencia de la **Universidad Simón Bolívar** de Barranquilla, a nivel local, regional, nacional e internacional.

1.2 ANTECEDENTES

Las políticas desarrolladas por el Gobierno Nacional en los últimos años PND (2018), para ampliar las posibilidades de acceso y la cobertura del sistema de educación superior, han traído nuevos retos a las instituciones, las cuales se han visto en la necesidad de reforzar sus acciones para promover no solo el ingreso de esos nuevos estudiantes, sino también la culminación exitosa de sus estudios. La graduación estudiantil es el reflejo del cumplimiento de los objetivos fundamentales de la educación superior. Este se logra a partir del desarrollo de procesos formativos y la implementación de estrategias de intervención pertinentes en cada etapa del ciclo del estudiante en la educación superior, desde su ingreso hasta su graduación. Sin embargo, el cumplimiento de este objetivo se ve afectado por los altos niveles de deserción de los estudiantes en el pregrado, una de las principales dificultades de los sistemas educativos del mundo. Los análisis realizados en el SPADIES (2020) Sistema para la Prevención de Deserción en Educación Superior, han permitido identificar que en los primeros cuatro semestres es el período en el que se concentra el

75 % del total de desertores. Los principales factores asociados a este fenómeno, en el caso colombiano, están relacionados con las bajas competencias académicas de entrada, las dificultades económicas de los estudiantes y los aspectos relacionados con la orientación socio ocupacional y la adaptación al ambiente universitario. Asimismo, de acuerdo con SPADIES, para 2015, la tasa de deserción anual en el nivel universitario se ubicó en el 9.3 % y la tasa de deserción por cohorte fue del 46,05 %; mientras que para el nivel técnico y tecnológico la deserción anual está por encima del 20 por ciento anual y el 55 % por cohorte 1, publicado en Ministerio de Educación Nacional Colombiano (2015) la Guía para la implementación para el modelo de permanencia y graduación estudiantil para instituciones de educación superior del año 2015. La tabla 1.1 a continuación nos muestra un panorama entre el 2014 y 2018 de acuerdo a la información de SPADIES (2020):

TABLA 1.1: Tabla con Porcentajes de deserción en Colombia

Tipo de Deserción	2014	2015	2016	2017	2018
Período(Anual)	12 %	9,3 %	9.03 %	10 %	9.5 %
Cohorte	54 %	46.05 %	45.09 %	45 %	44,6 %

1.3 DEFINICIÓN DEL PROBLEMA CIENTÍFICO

Como puede observarse en la discusión de antecedentes en la sección anterior, aunque se reconoce la necesidad de disminuir los índices de deserción escolar, a pesar de la necesidad reconocida ampliamente y del volumen de investigación científica realizada en el tema, no existen trabajos que aporten una metodología flexible, reproducible y con base científica que sustente el desarrollo de políticas de gestión y mitigación de la deserción escolar. Precisamente ese es el problema que se aborda en este proyecto de tesis, que responde a las siguientes preguntas de investigación:

1. ¿Es posible una metodología, desde una perspectiva científica de apoyo a la decisión para desarrollar políticas o estrategias de intervención efectivas orientadas a disminuir la deserción escolar en planteles de educación superior?
2. ¿De qué manera detectar a tiempo desertores en potencia para brindarle la clase de apoyo que requieran (académico, económico, psicológico, etc.) y así potenciar sus posibilidades de permanecer en la institución y continuar con sus estudios hasta graduarse?

1.4 HIPOTESIS

La atención por parte de las instituciones de educación superior a la permanencia estudiantil, busca cada vez integrar mas variables incidentes en el fenómeno de la deserción. En un marco ideal, estas variables deben estar asociadas a las dimensiones fundamentales que impactan en la deserción (Académicas, Institucionales, Económicas e Individuales o Antropológicas), siendo un reto la interpretación y evaluación del impacto de las variables en la permanencia del estudiante en una institución. Por otra parte, es común que las instituciones de educación superior cuenten con abundantes datos de las trayectorias académicas de sus estudiantes, así como de los mecanismos de acompañamiento académico, apoyos económicos, datos personales y los relacionados con el estado económico de los estudiantes. Datos que constituyen una invaluable fuente de información para guiar la toma de decisiones basadas en datos en el desarrollo de políticas y estrategias para disminuir los índices de deserción. Sin embargo, un reto adicional que enfrentan las instituciones es la interpretación de estos datos y sus consecuencias en la deserción, ya que por lo general la interpretación del impacto o la evaluación de las variables que se derivan de los datos tienen un fuerte componente subjetivo ya que debe estar orientadas con los objetivos de la institución.

Por ello, en este trabajo, se busca la aplicación del apoyo a la decisión mul-

ticriterio que ofrezcan un marco conceptual adecuado para desarrollar estrategias científicamente fundamentadas para mitigar las tasas de deserción escolar en la educación superior y la posible mejora en la calidad académica.

1.5 OBJETIVOS

Profundizar en el estudio del problema de deserción estudiantil desde un marco metodológico basado en la ciencia de datos y el apoyo a la decisión multicriterio, para establecer los factores principales que inciden en la deserción y contribuir al éxito estudiantil.

1.6 OBJETIVOS ESPECÍFICOS

- Desarrollar una metodología basada en la ciencia de datos, el análisis de la decisión multicriterio para establecer los factores principales que inciden en la deserción estudiantil.
- Mostrar la validez de la metodología, aplicándola a un caso de estudio real.

1.7 JUSTIFICACIÓN

Mitigar la deserción estudiantil es uno de los objetivos principales en las instituciones de educación superior. Sin embargo, a pesar de definir factores y variables incidentes en la permanencia (académicos, individuales, institucionales, económicos), las instituciones no cuentan con un modelo de gestión a la permanencia donde se analizan los datos como un conjunto y permita tomar decisiones frente a la implementación de estrategias de acompañamiento efectivas y eficaces, desconociendo el panorama real de lo que ocurre en los diferentes escenarios del estudiante, en

los servicios ofrecidos o de los que se carecen para atención a la permanencia. Es por ello, que se hace importante la definición de una metodología para la gestión de permanencia estudiantil de educación superior basado en el apoyo a la decisión multicriterio en búsqueda de la graduación oportuna de los estudiantes.

1.8 NOVEDAD CIENTÍFICA

Se presenta una metodología basada en la ciencia de datos, el apoyo a la decisión multicriterio para la mitigación de la deserción estudiantil en la educación superior y contribuir al éxito estudiantil.

1.9 ESTRUCTURA DEL DOCUMENTO

El presente documento está organizado de la siguiente manera: en el Capítulo 2, se presenta el marco teórico abarcando los temas que están relacionados con el problema y la perspectiva desde la cual se aborda el problema. En dicho capítulo se exponen temas como la ciencia de datos, la teoría de apoyo a la decisión multicriterio, deserción estudiantil, permanencia estudiantil entre otros. También, se presenta el estado del arte y se expone un análisis crítico de trabajos similares.

En el Capítulo 3 se plantea detalladamente el problema, se presenta el enfoque metodológico aplicado y los supuestos considerados. En el Capítulo 4 se describe la metodología general que se sigue para resolver el problema aquí abordado, en el capítulo 5 se desarrolla la aplicación de la metodología a un caso real de estudio. Finalmente, en el Capítulo 6 se presentan las conclusiones y recomendaciones para trabajo futuro.

CAPÍTULO 2

MARCO TEÓRICO Y ESTADO DEL ARTE

2.1 DESERCIÓN ESTUDIANTEL

2.1.1 ABANDONO ESCOLAR / DESERCIÓN

La deserción estudiantil universitaria, de acuerdo con Corena Gutierrez *et al.* (2012) es un fenómeno multicausal y multifactorial de las instituciones de educación superior, que causa impactos de importancia en todos aquellos actores que hacen parte del sistema, es decir, sobre los individuos, las instituciones de educación superior y el Estado. Se ha identificado que una de las problemáticas que más afectan las Instituciones de Educación Superior (IES) es el abandono de los estudiantes en diferentes momentos de su carrera técnica, tecnológica o profesional, viéndose afectado el índice de permanencia en los estudiantes, a pesar de diferentes estrategias de acompañamiento a los mismos, debido a la diversidad de nuestros estudiantes desde lo étnico y cultural, hasta el cambio generacional incidiendo en la forma de pensar, actuar de los mismos, teniendo que enfrentar a las ies a complejidades inéditas tal y como lo señala Strauss *et al.* (1991). Sin embargo, el problema de la deserción en

las Instituciones de Educación Superior, viene siendo estudiando desde hace tiempo atrás, donde se comenzaron a relacionar la influencia en la deserción de factores externos al individuo, adicionales a los psicológicos según Watson *et al.* (2004). Dicho autor logró identificar seis predictores de la deserción estudiantil, enmarcados en dos primeros años de estudio universitario norteamericano, clasificados de acuerdo a lo académico y lo personal, donde en lo académico se considera la integración Académica del estudiante en la IES, el promedio de notas en cada semestre y la calidad de la carrera; y dentro de lo personal el estado Socioeconómico, Género y la Integración Social del mismo.

2.2 CIENCIA DE DATOS

El concepto de la ciencia de datos es definida por Tukey (1962) como un conjunto de procedimientos para analizar datos, técnicas para interpretar los resultados de dichos procedimientos, formas de planificar la recopilación de datos para hacer su análisis más fácil, más preciso o acertado, y toda la maquinaria y los resultados de las estadísticas matemáticas que se aplican al análisis de datos, con el objetivo de aprender cada vez más acerca de los datos y sus modelos, para lograr una mejor precisión en cuanto al conocimiento del negocio.

De acuerdo con lo expresado por Moreno-Salinas (2017) hoy más que nunca estamos más conectados con personas y dispositivos, así como el acceso a más redes y servicios, consumiendo de esa manera mayor cantidad de datos e información, por lo cual se requiere contar con las habilidades, conocimientos, experiencias y técnicas de los científicos de datos para procesar, analizar y visualizar de formas más inteligentes los datos en información, promoviendo así, más y mejores conocimientos de nuestra realidad en sus contextos, en especial con los datos de la organización.

El padre del “management” **Peter Drucker** (Drucker (2004)) reconoció que la sociedad postcapitalista es una sociedad basada en el conocimiento, donde el centro

de la producción de la riqueza es el saber y no el capital, por lo cual afirma que en esta época, los protagonistas serán los **trabajadores del conocimiento**, con las capacidades, las habilidades, el pensamiento creativo y la tecnología para procesar, analizar y visualizar las grandes bases de datos, a los que se les conoce actualmente como **Científicos de datos**, y son quienes combinan estadística, matemáticas, programación y solución de problemas, con la captura datos de forma ingeniosa y la capacidad de mirar las cosas de manera diferente (encontrar patrones), además de hacer las actividades propias de limpieza, preparación e integración de datos (Monnappa, 2017).

En la siguiente imagen 2.1 se describen, las habilidades, los conocimientos y la experiencia que debe poseer el científico de datos, así como una muestra del software especializado y técnicas que existe por sus áreas de desarrollo para esta década:

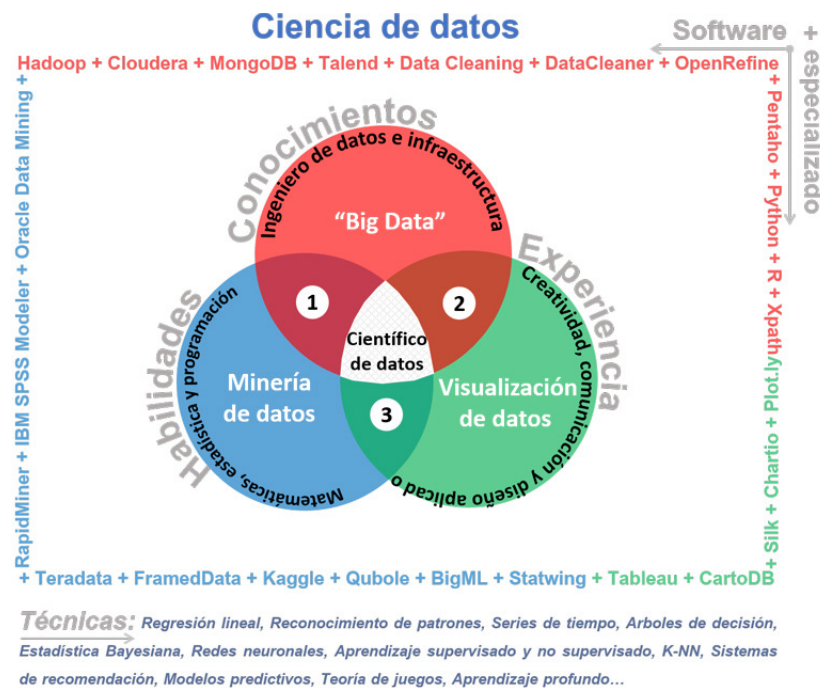


FIGURA 2.1: Científico de Datos y su entorno. Moreno-Salinas

Por lo tanto, desde ese análisis, el científico de datos debe dar solución a las preguntas relacionadas con la forma de almacenamiento de datos, de qué manera

analizar y obtener el valor de los datos, cómo expresar lo que los datos quieren decir (visualización y comunicación). Sin embargo, VanDer-Aalst (2014) sugiere que el científico de datos puede llegar a resolver las preguntas ¿qué paso?, ¿por qué paso?, ¿qué sucederá?, ¿qué es lo mejor que puede pasar?, preguntas relacionadas con la minería de datos.

2.2.1 METODOLOGÍA FUNDAMENTAL PARA LA CIENCIA DE DATOS IBM

La empresa International Business Machines (IBM), a través del Dr John B. Rollins (2015), propone una metodología fundamental para la ciencia de datos, capaz de proporcionar una estrategia de orientación, que sea independiente de las tecnologías, los volúmenes de datos o los enfoques involucrados, que consta de 10 etapas, como lo muestra la imagen 2.2

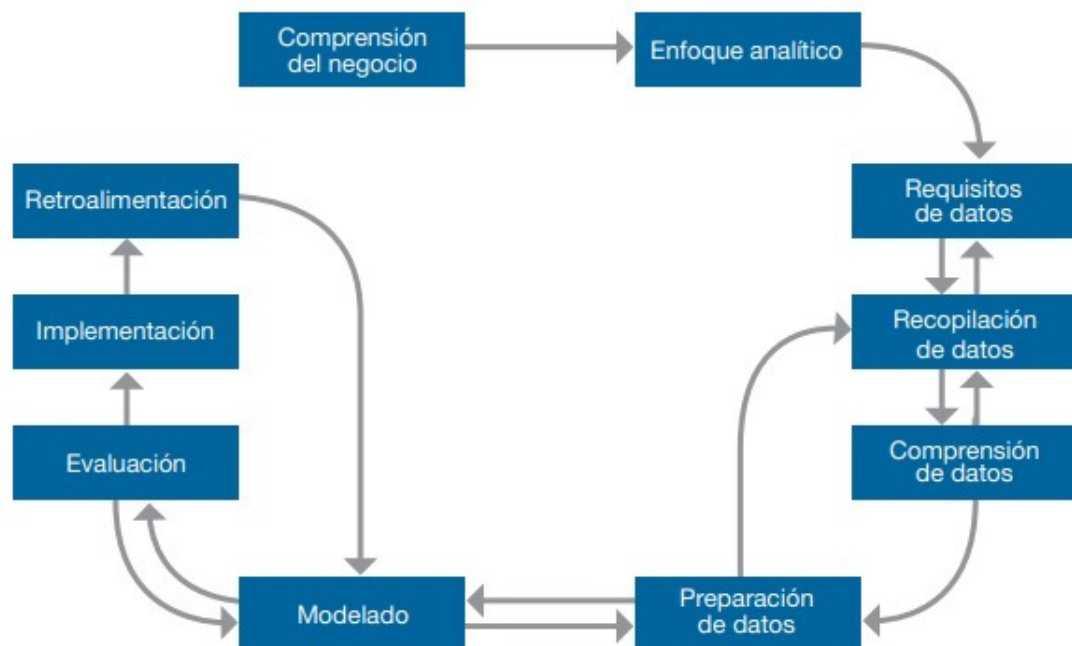


FIGURA 2.2: Metodología Fundamental para la ciencia de datos(IBM)Rollins

Las etapas que propone el modelo son descritas a continuación:

1. **Comprensión del Negocio:** Una de las etapas mas importantes, ya que requiere del conocimiento de las características que inciden en el modelo del negocios donde se aplicará la analítica de datos. El entender el significado de los datos para los procesos del negocio, toman gran relevancia al momento de definir los objetivos y requisitos de solución de los proyectos.
2. **Enfoque Analítico:** Aquí se define que es lo que se quiere como resultado, para que la organización obtenga de la mejor manera la solución al problema identificado en la etapa uno (1). Esta etapa implica expresar el problema bajo el contexto de las técnicas estadísticas y de aprendizaje automático, para que la organización pueda identificar las más adecuadas para el tipo de resultado deseado, por ejemplo, el enfoque puede definirse como la construcción, las pruebas y la implementación de un modelo de clasificación en caso que se requiera predecir una respuesta de acuerdo al análisis de los datos y su correlación Rollins (2015).
3. **Enfoque de datos:** De acuerdo al enfoque analítico seleccionado, entonces, se deben identificar los datos que se requieren y su disponibilidad en la organización. En esta etapa, se busca que se defina cuales datos se requieren para poder solucionar.
4. **Recopilación de datos:** En esta etapa, se identifican y reúnen los recursos de datos disponibles (estructurados, no estructurados y semiestructurados) y relevantes para el dominio del problema, así como definir la muestra y cantidad de datos a utilizar. Es posible que no se cuente con todos los datos que se definieron en el enfoque de datos y en ese caso, hay que hacer una reestructuración de los datos necesitados para el análisis. En este tipo de metodología a mayor datos, mayor posibilidad de generar un modelo predictivo con una representación de eventos raros o fallo en el sistema.
5. **Comprensión de datos:** En cuanto termina el proceso de recopilación de datos

iniciales, se inicia el proceso de las estadísticas descriptivas y aplicación de técnicas de visualización para poder encontrar el significado de los datos, al igual que evaluar su calidad y descubrir insights iniciales sobre los mismos.

6. Preparación de los datos: En esta fase se inicia el proceso de construcción del conjunto de datos. Entre las actividades que se realizan en la etapa encontramos la limpieza de datos, combinar los datos de múltiples fuentes y la transformación de datos en variables más útiles. Se aplica la denominada **ingeniería de características**, para crear variables explicativas o indicadores combinando el conocimiento de dominio y las variables estructuradas existentes. Beltran (2016) considera que entre las acciones a realizar en la fase de preprocesamiento de datos, con respecto a datos anómalos o con errores están: ignorarlos, filtrarlos (eliminando o reemplazando una columna), filtrado de filas, reemplazo de valores (por el valor “nulo”, máximos, mínimos o medias) o discretización (transformar los valores continuos por discretos). Esta fase suele ser la más larga del proceso.
7. Modelado: Desde el conjunto de datos inicial, se procede al desarrollo de los modelos predictivos o descriptivos, dependiendo de los definidos en el enfoque analítico. Normalmente, para los modelos predictivos se tienen en cuenta datos históricos en los que se conoce el resultado del interés (Conjunto de capacitación). En la actualidad esta etapa es muy iterativa, ya que se puede contar con insight intermedios que permiten la preparación de datos y ajustes en la especificación del modelo. Se pueden aplicar múltiples algoritmos para llegar a encontrar el mejor modelo para las variables disponibles.
8. Evaluación: En esta etapa el científico de datos, debe evaluar el modelo de la etapa anterior, desde su calidad y pertinencia y coherencia al problema presentado. Esta evaluación implica el cálculo de algunas medidas de diagnóstico y otros resultados, ya sea utilizando un conjunto de datos de prueba, diferente al conjunto de capacitación utilizado para el modelo, considerando, la misma distribución de probabilidad y tiene un resultado conocido.

9. Implementación: Luego de la aprobación por directores del negocio del modelo, se aplica en el entorno de pruebas comparable o de producción de la empresa, con datos actualizados, normalmente durante un tiempo determinado, hasta que se haya podido evaluar su rendimiento de manera completa. Este paso involucra a grupos, habilidades y tecnologías adicionales dentro de la empresa.
10. Retroalimentación: En esta fase la empresa obtiene una retroalimentación sobre el rendimiento del modelo al igual que su impacto en el entorno implementado, ya sea en forma de porcentajes de respuesta o de otra manera. Desde esta retroalimentación el científico de datos, puede ajustar el modelo para poder mejorar su precisión y utilidad para obtener mejores resultados, así como también implementar la automatización de ciertos pasos de la evaluación del modelo.

La metodología propone una constante actualización de los modelos, manteniendo abierta la mejora a través de ajustes y reimplementaciones todo el tiempo

2.2.2 CRISP - DM

La metodología Cross Industry Standard Process for Data Mining mas conocida como CRISP-DM, presenta ocho (8) fases resumidas en seis (6) como lo muestra la figura 2.3

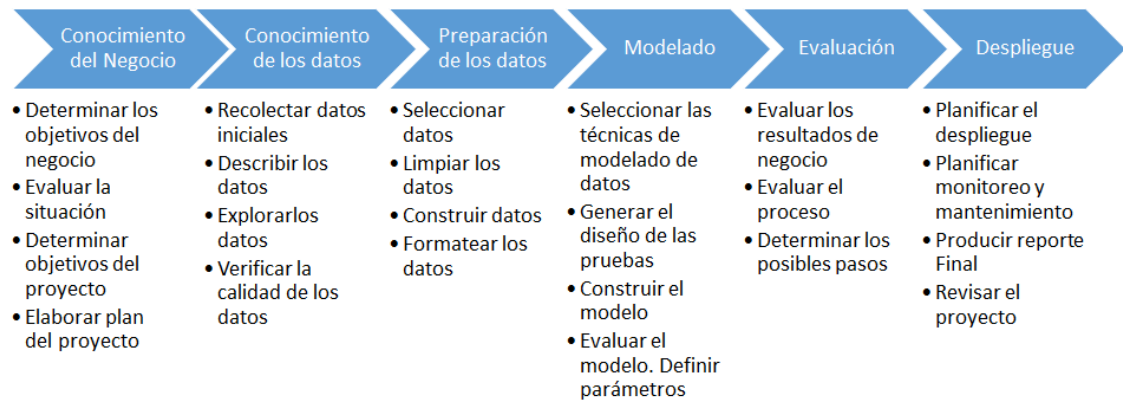


FIGURA 2.3: Resumen Metodología. Propio

2.2.3 APRENDIZAJE AUTOMATIZADO

Para las investigación que incluyen el procesamiento de los datos existen dos enfoques: el enfoque orientado a los datos (Data-Driven Approaches - DDA) y el enfoque orientado al conocimiento (Knowledge-Driven Approaches – KDA). El DDA incluye las categorías de: análisis de umbral (Threshold Analysis), métodos de aprendizaje automático (Machine Learning Methods – MLM), métodos de regresión (Regression Methods) y técnicas de correspondencia de curvas (Curve Matching Techniques). Según Camargo García (2020) como resultado del exhaustivo análisis de la literatura científica, se evidencia que la categoría MLM es de las más utilizadas para procesos de clasificación para la predicción. Machine learning de acuerdo con Hurwitz y Kirsch (2018) es una forma de la Inteligencia Artificial que permite a un sistema aprender de los datos en lugar de aprender mediante la programación explícita, que puede lograr la generación de modelos que se definen como la salida de información que se genera cuando entrena un algoritmo de machine learning con datos (algoritmo predictivo) y creará un modelo predictivo obteniendo un pronóstico basado en los datos que entrenaron al modelo. El Aprendizaje Automático (MLM) contempla, entre otros, los siguientes métodos: clasificadores bayesianos (Bayesian Classifiers - BC), clasificadores basados en instancias (Instance Based Classifiers), máquinas de soporte vectorial (Support-Vector Machine – SVM), árboles de decisión (Decision

Trees – DT), redes neuronales artificiales (Artificial Neural Networks), lógica difusa (Fuzzy Logic) y modelos de markov (Markov Models), etc.

2.2.3.1 ÁRBOLES DE DECISIÓN

Quinlan (1986) define los Árboles de Decisión como métodos de aproximación a las funciones de valor discreto que es robusto a los datos ruidosos y capaz de aprender expresiones disyuntivas. Se construyen comenzando con la raíz del árbol y procediendo hasta sus hojas. Los métodos de árbol de decisión están diseñados para seleccionar un conjunto de variables de predicción y dividir sucesivamente un conjunto de datos en subgrupos a fin de mejorar la predicción (clasificación) de una variable objetivo (dependiente), los cuales pueden, representarse como conjuntos de reglas Si y o entonces para una mejor comprensión. Los datos manejados en los árboles son representados con tres elementos:

1. Atributos/Predictores
2. Valores
3. Clases

Por otro lado, existen dos tipos de árboles de decisión:

- De clasificación
- De regresión

En los árboles de clasificación, según Quinlan (1986), cuando las variables son discretas, predice el valor de una variable categórica mediante la creación de un modelo basado en uno o más atributos.

2.2.3.2 MÁQUINAS DE SOPORTE VECTORIAL

Las máquinas de soporte vectorial – (SVM) nace en los años 90, fundamentados en la teoría del aprendizaje estadístico, fueron planteadas por Boser B. E. (1992). Se pensaron con el objetivo de resolver dudas en la clasificación binaria, las cuales con el pasar los años se han utilizado para optimización de agrupamiento, multclasificación, regresión e igual modo, en especial en las áreas de identificación de imágenes, clasificación de texto e hipertexto, procesamiento de lenguaje natural, estudio de series temporales entre otras.

La máquina de soporte vectorial considera que los vectores de entrada son mapeados no linealmente a un ámbito de atributos de muy alta extensión. En el área de los atributos se construye una extensión de decisión lineal y las propiedades especiales de la extensión de decisión aseguran una alta capacidad de generalización de la máquina de aprendizaje. En Han (2012) indican que una SMV utiliza un mapeo no lineal para transformar los datos originales de la formación en una dimensión superior. Dentro de esta nueva dimensión, busca el hiperplano de separación óptimo lineal. Los SVMs pueden ser usados para la predicción numérica, así como para la clasificación. El objetivo del algoritmo de la máquina de soporte vectorial es encontrar un hiperplano en un espacio de dimensiones N (N - el número de características) que clasifique claramente los puntos de datos. Para separar las dos clases de puntos de datos, hay muchos hiperplanos posibles que podrían ser elegidos y se pretende encontrar un plano que tenga el margen máximo, es decir, la distancia máxima entre los puntos de datos de ambas clases y proporcionar algún refuerzo para que los futuros puntos de datos puedan ser clasificados con más confianza

- Fortalezas de SVM De acuerdo con Betancourt (2005) se destacan las siguientes fortalezas de las máquinas de soporte vectorial:

- Entrenamiento fácil
- No existe un óptimo local

- Escalamiento bueno para datos en espacio con dimensiones altas
 - La responsabilidad entre el error y la complejidad del clasificador puede ser contrastado claramente.
 - Cadenas de caracteres y arboles pueden ser usados como datos no tradicionales como entrada a la SVM, en vez de vectores de características
- Debilidades de SVM: Es necesario contar con una buena función kernel, en otras palabras, se requieren procedimientos eficientes para recibir los parámetros de inicialización de la SVM

2.2.3.3 CLASIFICADORES BAYESIANOS

La clasificación es una labor sencilla en el estudio de datos y la identificación de patrones que necesita la construcción de un clasificador, en otros términos, una función que concede una etiqueta de clase a las instancias descritas por un conjunto de atributos (Friedman et al., 1997). Las redes bayesianas se consideran una posibilidad a los sistemas expertos clásicos dirigidos a la toma de decisiones y al pronóstico bajo indecisión en lo referente a probabilísticos (Picard et al., 2004). Así mismo, en (García et al., 2006) plantean que las redes de Bayes son herramientas estadísticas orientadas a la modelación causal. Una red de Bayes tiene dos componentes principales: una dimensión cualitativa basada en la teoría de los grafos y una cuantitativa basada en la teoría de la probabilidad. Una red bayesiana está compuesta por nodos de casualidad, unidos con otros nodos por arcos. Estos arcos están dirigidos: el arco va de un nodo padre a un nodo hijo. El estado en el que se encuentra un nodo padre no se ve afectado por el estado en el que se encuentran sus hijos. Pero la probabilidad de cada uno de los estados en los que puede estar un nodo hijo depende del producto cartesiano de los estados actuales de todos sus nodos padres según Edwards y Fasolo (2001) El clasificador de red Bayesiana aprende de los datos de preparación la probabilidad eventual de cada atributo A_i dada la etiqueta de clase C . La clasificación se hace entonces adaptando la regla de Bayes para evaluar la probabilidad de C

dada la instancia específica de A_1, \dots, A_n y luego prediciendo la clase con la probabilidad posterior más alta (Friedman, 1997). Los clasificadores tales como, Árboles de Decisión, Maquinas de Soporte Vectorial y Redes Bayesianas, son ampliamente utilizados para la predicción de datos de diferentes estudios tales como los de los autores Hasbun *et al.* (2016), Kalles y Pierrakeas (2006), Pimpa (2013) entre otros.

2.3 TEORIA DE APOYO A LA DECISIÓN

2.3.1 EL MODELO DE RACIONALIDAD LIMITADA

La teoría de la racionalidad limitada de Simon (1960) plantea que las personas tomamos decisiones de forma parcialmente irracional a causa de nuestras limitaciones cognitivas, de información y de tiempo.

Este modelo surgió como reacción a las teorías de la racionalidad, muy populares en las ciencias políticas y económicas, que proponen que los humanos somos seres racionales que decidimos cuál es la solución óptima para cada problema utilizando toda la información disponible.

No obstante, según Simon y los autores que lo sucedieron, es muy difícil tomar decisiones totalmente racionales porque nuestros recursos para procesar la información son limitados, en especial cuando los problemas son complejos, como suele ocurrir en la vida cotidiana. Frente a la idea clásica del "hombre económico", Simon promovió la del 'hombre administrativo', incapaz de captar la complejidad del mundo y la interrelación entre sus elementos.

El modelo de racionalidad limitada afirma que las personas usamos los heurísticos a la hora de encontrar soluciones. Los heurísticos se definen como reglas generales y sencillas que usamos para resolver problemas; aunque pueden ser útiles en muchos casos, en otros producen sesgos cognitivos, es decir, desviaciones sistemáticas en el

razonamiento.

El proceso de toma de decisiones sigue las fases siguientes de acuerdo a Simon (1960) como se observa en la la siguiente gráfica 2.4 :

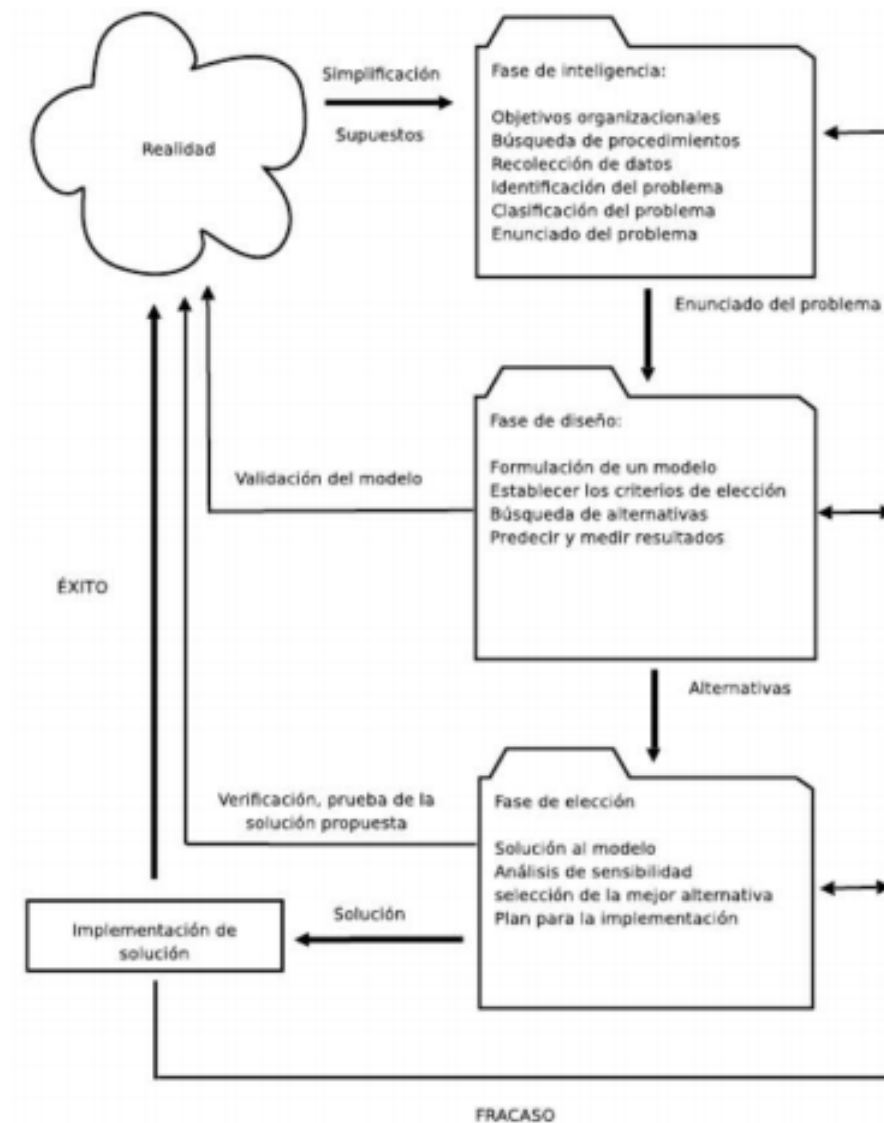


FIGURA 2.4: Descripción Metodología Simon

2.3.1.1 TOMA DE DECISIONES BAJO CRITERIOS MÚLTIPLES

Un tomador de decisiones frecuentemente trata con problemas que involucran múltiples criterios, usualmente en conflicto. El proceso de toma de decisiones multi-criterio (Multiple Criteria Decision Making (MCDM)) se refiere a tomar decisiones sobre un conjunto de alternativas que se caracterizan por múltiples atributos o criterios en conflicto (Yoon y Hwang, 1995). Los problemas de MCDM se caracterizan por tener:

- Tomador de decisiones (Decision Maker o DM en inglés): la persona o personas directamente responsables de las consecuencias de la toma de decisiones en el problema o situación abordada.
- Alternativas: un número finito de alternativas (opciones, acciones o candidatos).
- Múltiples criterios: cada problema tiene múltiples atributos o criterios y la cantidad dependerá de la naturaleza del problema.
- Escalas de preferencia (representan consecuencias de la realización de criterios en cada alternativa): se tienen diferentes medidas de medición por criterio.
- Atributos de peso o importancia relativa: la mayoría de los métodos de MCDM requieren información de la importancia de cada criterio, proporcionada ya sea por una escala ordinal o cardinal. Los pesos se asignan por el tomador de decisiones o pueden ser desarrollados por métodos (MAUT – Multi Attribute Utility Theory (Teoría de la Utilidad de Múltiples Atributos), AHP – Analytic Hierarchy Process (Proceso Analítico Jerárquico), Fuzzy Theory (Teoría de Conjuntos Difusos), CBR - Case Based Reasoning (Razonamiento Basado en Casos), DEA – Data Envelopment Analysis (Análisis Envolvente de Datos), SMART – Simple Multi Attribute Rating Technique (Técnica Simple de Calificación de Múltiples Atributos), Goal Programming (Programación de objetivos), ELECTRE, PROMETHEE, SAW – Simple Additive Weighting

(Ponderación Aditiva Simple). TOPSIS – Technique for Order of Preference by Similarity to Ideal Solution,

- Matriz de decisión: un problema MCDM puede ser expresado en una matriz, las columnas son los criterios y las filas corresponden a las alternativas, las celdas se corresponden a la consecuencia de un atributo o criterio para una alternativa.

2.3.1.2 CLASE DE PROBLEMAS DE TOMA DE DECISIONES

Según ? los problemas de decisión en el marco de las organizaciones, de acuerdo con Yevseyeva (2007) se clasifican en:

1. Problema de jerarquización de alternativas: cuando es necesario ordenar las alternativas del conjunto dado según las preferencias del DM. Un ejemplo de tal problema es la asignación de los participantes de una competencia entre los mejores y los peores.
2. Problema de selección de alternativas: cuando durante el proceso de ayuda a la decisión, el DM debe encontrar el menor número de las mejores alternativas posibles. Por ejemplo, puede ser necesario elegir solo una universidad para estudiar, o elegir a una persona en particular para el puesto de gerente.
3. Problema de ordenamiento/clasificación de alternativas: cuando las alternativas deben asignarse en una de varias clases del conjunto de posibles. Este tipo de problema aparece, por ejemplo, en la tarea de diagnóstico (por ejemplo, médica, técnica y financiera). Por ejemplo, un médico define una enfermedad según los síntomas del paciente. Otro ejemplo de la vida cotidiana es que podemos clasificar las tareas "pendientes", para el día siguiente en cuatro clases, como "debe hacer", "deseo hacer"; y "puedo hacer".

En este trabajo nos centraremos en el problema de clasificación. De acuerdo con

lo investigado, existen diferentes tipos de métodos para resolver problemas de clasificación con la ayuda de enfoques MCDA. En muchas situaciones, el tomador de decisiones se enfrenta al problema de agrupar alternativas en clases de homogeneidad: a esto lo llamamos problema de clasificación. Por lo general, una tarea de calificaciones aparece en los diagnósticos, cuando el tomador de decisiones asigna cada alternativa del conjunto de alternativas dadas a una o varias clases.

Otra clasificación de los problemas de decisión se realiza en base a la naturaleza de las consecuencias asociadas a cada alternativa:

1. Decisión bajo certeza: Cuando la consecuencia a cada alternativa esta perfectamente identificada al igual que su ocurrencia.
2. Decisión bajo incertidumbre: Cuando se conocen las consecuencias asociadas a cada alternativa, pero la ocurrencia de estas se define mediante variables aleatoria con distribución conocida.
3. Decisión bajo o estricta incertidumbre: Cuando se conocen las consecuencias asociadas a cada alternativa, pero no se tiene ninguna información sobre su ocurrencia, se desconoce incluso si la ocurrencia puede modelarse mediante variables probabilísticas.

2.3.1.3 ELECTRE

La manera de seleccionar a una alternativa entre varias, en los problemas de toma de decisiones, es comparándolas con base a ciertos atributos o criterios de evaluación. Para cada criterio se establece un peso de importancia (w). El método ELECTRE permite evaluar las ventajas y desventajas entre las alternativas en cada criterio y jerarquizarlas en un orden de preferencias, de la mejor a la peor. Este método fue creado por Roy en 1969 y desde entonces se ha modificado varias veces. ELECTRE se forma como acrónimo de la frase en francés, **EL**imination **Et** **Choix**

TRaduisant la realitE y es un método comparativo de opciones o alternativas que se basa en operaciones efectuadas en tablas o matrices, a partir de las calificaciones en los criterios de cada alternativa. Se conocen cinco versiones de la herramienta: Electre I, Electre II, Electre III, Electre IV y el mas reciente Electre TRI.

ELECTRE TRI Está basado en el concepto de relaciones de subordinación. Una relación de subordinación, es una relación binaria S , definida dentro de un conjunto de acciones A , tal que aSb , se lee a subordina a b, si dadas las preferencias del decisor, hay suficientes argumentos para decidir que a es tan buena como b, siempre y cuando no haya razones de peso que refuten esta afirmación Vincke (1992)). Este método realiza dos procedimientos de clasificación, uno OPTIMISTA y el otro PESIMISTA, los cuales, en función a las puntuaciones obtenidas en cada alternativa por criterio, buscan clasificar cada alternativa en relación a un conjunto de categorías predefinidas, donde cada categoría es limitada por dos perfiles. La decisión de cual clasificación utilizar dependerá del decisor. Además el ELECTRE TRI trabaja con índices tales como:

- Índice de concordancia parcial y global: el cual expresa que tan buena es una alternativa en relación con un perfil determinado. Es parcial si se considera un criterio y es Global si se consideran todos.
- Índice de Discordancia parcial y global: expresa hasta que punto se discrepa con la afirmación de que tan buena es una alternativa en relación a un determinado perfil. Al igual que el anterior, si se considera un criterio es parcial y si se consideran todos es global.
- Índice de credibilidad: expresa que tanto es creíble la afirmación de que una alternativa es preferible a un determinado perfil en todos los criterios.
- Límite de corte λ : Determinado por el decisor, varía entre $0.5 \prec \lambda \prec 1.0$. Es el menor valor del índice de credibilidad compatible con la afirmación de que una alternativa subordina a un perfil. Los valores de λ mas cerca de 1, vuelven los resultados mas rigurosos.

- Límite de preferencia (p): Representa la mínima diferencia aceptada entre el puntaje de una alternativa y un perfil para un criterio cualquiera, de forma que se considere que existe preferencia entre ellos.
- Límite de indiferencia(q): Representa la máxima diferencia aceptada entre el puntaje de una alternativa y un perfil para un criterio cualquiera, de forma que se considere que no existe preferencia entre ellos.
- Límite de veto (v): Representa la mínima diferencia entre las puntuaciones de una alternativa y un perfil, para que se invalide la afirmación de que la alternativa sea preferible al perfil.

2.3.2 CONJUNTOS APROXIMADOS. ROUGHT SET

La teoría de conjuntos aproximados se define, de acuerdo con Pawlak (1998), como una herramienta para describir los atributos entre las dependencias para evaluar su importancia y tratar con datos inconsistentes en los procesos de decisión. En los procesos de decisión es necesario un modelo de preferencia en la teoría de conjuntos aproximados dicho modelo esta formado por reglas de decisión de la forma: "si x entonces y". La teoría de conjuntos aproximados se aplica cuando las clases en las que han de clasificarse los objetos son imprecisas, pero pueden aproximarse mediante conjuntos precisos. La idea básica del razonamiento basado en conjuntos aproximados sobre los datos es más de carácter inductivo que deductivo. Se logra una relación entre una implicación en la lógica deductiva y una regla de decisión en el enfoque del conjunto aproximado. El razonamiento basado en conjuntos aproximados se superpone con el razonamiento deductivo, inductivo y de sentido común, sin embargo, tiene sus propias características específicas y puede considerarse por derecho propio. Esta teoría según Chin K.S. ((2003) se considera como una de las cinco áreas claves y no tradicionales de la Inteligencia Artificial y de la Teoría de la Información Incompleta, pues constituye una herramienta muy útil para el manejo

de la información no completa o imprecisa.

2.3.2.1 TABLAS DE DECISIÓN

En la teoría de conjuntos aproximados, la base es un conjunto de datos (información) sobre algunos objetos de interés organizados en forma de tabla denominada, de acuerdo a Pawlak (1998), sistema de información o tabla de información. Un ejemplo clásico y sencillo de una tabla de información 2.1, la cual es presentada a continuación por Pawlak (1998), la cual describe seis autos en términos de sus (atributos) características tales como: consumo de combustible (CC), calidad percibida (Q); precio de venta (P) y comerciabilidad (Co):

Objeto	CC	Q	P	Co
1	Alto	Regular	Med	Mala
2	Muy alto	Buena	Med	Mala
3	Alto	Buena	Bajo	Mala
4	Med	Regular	Med	Buena
5	Muy Alto	Regular	Bajo	Mala
6	Alto	Buena	Bajo	Buena

TABLA 2.1: Tabla Atributos del automóvil

El problema puede estar en relacionar la comerciabilidad entre las diferentes características seleccionadas, para poder dar cuenta de lo que influye en la aceptación de los autos en el mercado. Al seleccionar atributos específicos para comparar, se crean las tablas de decisión, dónde cada fila de una tabla de decisiones determina una regla de decisión, la cual especifica decisiones (acciones) que deben tomarse cuando se satisfacen las condiciones señaladas por los atributos de condición. Por ejemplo, en la Tabla 2.1, la condición (CC, alto), (Q, regular), (P, med) determina únicamente la decisión (Co, mala). Las reglas de decisión 3 y 6 de la Tabla 2.1 tienen las mismas condiciones pero diferentes decisiones. Tales reglas se denominan

inconsistentes (no deterministas, conflictivas, posibles). Si no fuese así, entonces las reglas se denominan consistentes (ciertas, deterministas, no conflictivas). Las tablas de decisión que contienen reglas de decisión inconsistentes se denominan inconsistentes (no deterministas) y en caso de las reglas consistentes, la tabla es coherente (determinista, etc.). El número de reglas consistentes para todas las reglas en una tabla de decisión se puede utilizar como factor de coherencia de la tabla de decisión y se indicará con $\gamma(C, D)$, donde C y D son atributos de condición y decisión respectivamente. Por lo tanto, si $\gamma(C, D) < 1$, la tabla de decisiones es consistente y si $\gamma(C, D) \neq 1$, la tabla de decisiones es inconsistente. Por ejemplo, para la Tabla 2.1 $\gamma(C, D) = 4/6$. En lo que sigue, los sistemas de información se denotarán por $S = (U, A)$; donde U es el universo, A es un conjunto de atributos, de modo que para cada $x \in U$ y $a \in A$, $a(x) \in V_a$ y V_a es el dominio (conjunto de valores) de a, de acuerdo con (Pawlak, 1998).

2.3.2.2 REGLAS DE DECISIÓN Y FACTOR DE CERTEZA

Las reglas de decisión normalmente se presentan como condiciones en el formato "Si.... entonces...". Por ejemplo, los condicionales de la tabla 2.1 son:

1. si (CC, alto) y (Q, regular) y (P, med) entonces (Co, mala)
2. si (CC, v. Alto) y (Q, bueno) y (P, medio) entonces (Co, mala),
3. si (CC, alto) y (Q, bueno) y (P, bajo) entonces (Co, mala),
4. si (CC, med.) Y (Q, regular) y (P, med.) Entonces (Co, buena),
5. si (CC, v. Alto) y (Q, regular) y (P, bajo.) Entonces (Co, mala),
6. si (CC, alto) y (Q, bueno) y (P, bajo) entonces (Co, buena),

En general, las reglas de decisión son condiciones construidas a partir de fórmulas elementales (nombre de atributo, valor de atributo) y combinadas mediante ope-

radores lógicos y, o e implicación. Sean Φ y Ψ fórmulas lógicas que representan condiciones y decisiones, respectivamente y $\Phi \rightarrow \Psi$ una regla de decisión, donde Φ_s denota el significado de Φ en el sistema S; es decir, el conjunto de todos los objetos que satisfacen Φ en S. Con cada regla de decisión $\Phi \rightarrow \Psi$ se asocia un número, llamado factor de certeza de la regla, y se define así:

$$\mu_s(\Phi, \Psi) = \frac{|\Phi_s \cap \Psi_s|}{|\Phi_s|},$$

dónde $|\Phi|$ denota la cardinalidad de Φ . Claro está $0 \leq \mu_s(\Phi, \Psi) \leq 1$; si la regla $\Phi \rightarrow \Psi$ es consistente entonces $\mu_s(\Phi, \Psi) = 1$, y para reglas inconsistentes $\mu_s(\Phi, \Psi) < 1$. Entonces, el factor de certeza para la regla de decisión 2 es 1 y para la regla de decisión 3 es 0,5. El factor de certeza se puede interpretar como una probabilidad condicional de una decisión Ψ dada la probabilidad de la condición Φ . La implicación en lógica deductiva se utiliza para sacar conclusiones a partir de premisas, mediante reglas de inferencia. Al razonar sobre las implicaciones de los datos hay reglas de decisión utilizadas para describir patrones en datos. En el enfoque de conjunto aproximado, las reglas de decisión también pueden trabajar de forma similar a la lógica deductiva y se puede considerar la siguiente fórmula:

$$\pi_S(\Psi) = \Sigma(\pi_S(\Phi) \cdot \mu_S(\Phi, \Psi)) = \Sigma \pi_S(\Phi \wedge \Psi),$$

donde Σ se toma sobre todas las condiciones Φ asociadas con la decisión correspondiente a Ψ , y $\pi_S(\Psi) = \frac{|\Phi_S|}{|U|}$. $\pi_S(\Phi)$ es una probabilidad de que la condición Φ se satisfaga en S. La fórmula muestra la relación entre la probabilidad de condiciones, el factor de certeza de una regla de decisión y la probabilidad de decisiones a tomar y permite calcular la probabilidad de que la decisión Ψ sea satisfecha en S, en términos de probabilidad de la condición Φ y probabilidad condicional de la regla de decisión $\Phi \Rightarrow \Psi$

2.3.2.3 APROXIMACIÓN DE CONJUNTOS

La teoría conjuntos aproximados es útil cuando las clases en las que han de clasificarse los objetos son imprecisas, pero sin embargo pueden aproximarse mediante conjuntos precisos. De manera intuitiva, un conjunto inexacto es un conjunto de objetos que, en general, no pueden ser caracterizados de manera precisa en terminos de la información disponible. Para el caso de los autos, de acuerdo a lo observado no hay una característica única para los que se venden bien (Co = mal), se conformarán dos conjuntos, llamados conjunto inferior y conjunto superior aproximados de automóviles que se venden bien. Cualquier subconjunto B de A determina una relación binaria I_B sobre U , la cual podría ser llamada relación de indiscernibilidad definiendola así: xI_By si y solo si $a(x) = a(y)$ para cada $a \in B$, donde $a(x)$ denota el valor del atributo a para el elemento x . I_B es una relación de equivalencia. La familia de toda equivalencia clases de I_B , es decir, la partición determinada por B , será denotada por U/I_B , o simplemente $U = B$; una clase de equivalencia de I_B , es decir, el bloque de la partición $U = B$; que contiene x se indicará con $B(x)$. La relación de indiscernibilidad se utilizará para definir conceptos básicos de teoría de conjuntos aproximada y se definen las siguientes dos operaciones sobre los conjuntos:

$$B_*(X) = \{x \in U : B(x) \subseteq X\}$$

$$B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\}$$

asignar a cada subconjunto X del universo U , dos conjuntos $B_*(X)$ y $B^*(X)$ llamados B inferior y B superior aproximado de X respectivamente. El conjunto:

$$BN_B(X) = B^*(X) - B_*(X)$$

se denominará región de límite B de X : Si la región de límite de X es el conjunto vacío, es decir, $BN_B(X) = \emptyset$ entonces el conjunto X es exacto con respecto al grupo B y en caso contrario, al ser inexacto o $BN_B(X) \neq \emptyset$ entonces el conjunto

es aproximado. Dichos conjuntos se denominan el reducto, el cual se define como el conjunto de criterios mas discriminantes. Al cociente entre la aproximación por debajo y la aproximación por encima se le denomina precisión.

La dependencia entre atributos es importante porque permite eliminar atributos redundantes. El reducto o conjunto minimo se define como el menor conjunto de atributos que mantiene la misma calidad de clasificación que el conjunto con todos los atributos y empleando la noción de reducto Pawlak (1998) podemos simplificar el conjunto de reglas de decisiones.

2.4 REVISIÓN DEL ESTADO DEL ARTE

2.4.1 DESERCIÓN O PERMANENCIA ESTUDIANTIL

Existe entonces una creciente literatura que apunta hacia múltiples factores. La propuesta que surge de la literatura es que la deserción o la permanencia dependen fuertemente de un exitoso acoplamiento entre la institución y el individuo. Este acoplamiento está mediado por factores individuales e institucionales. Dependiendo del enfoque, el principal problema puede ser el estudiante o la institución. A partir de aquí se empiezan a revisar ciertos modelos de deserción que nos resaltan de acuerdo a cada autor la importancia de cada variable dentro de cada factor a intervenir. Sin embargo, los modelos pueden ser tomados en su totalidad para aplicar en las IES, o simplemente hacer una mezcla de acuerdo a las características propias de la institución y sus estudiantes. Así como para Tinto (1975), la integración del estudiante es crucial, y ocurre cuando el estudiante se adapta a los valores, normas y prácticas universitarias y considera que esta integración depende de condiciones institucionales como el contacto con otros miembros de la comunidad universitaria o los métodos de enseñanza. Así mismo, las observaciones de Watson *et al.* (2004) acerca de la importancia de la cultura institucional para generar un acoplamiento

entre individuo e institución, y los aspectos como la flexibilidad curricular, la organización administrativa, la infraestructura y el ambiente académico y social, parecen influir sobre la permanencia. Así, una atención apropiada al estudiante, que considera aspectos académicos, familiares, sociales y económicos, resultaría importante para los jóvenes en el momento de decidir sobre su permanencia o abandono. Rausch y Hamilton (2006) revisaron estos factores de manera más cuantitativa al analizar estudiantes que habían decidido abandonar la universidad. Los factores mas destacados que encontraron fueron la dificultad para socializar y adaptarse al ambiente universitario, la sensación de aislamiento social y académico, así como la apreciación de una falta de integración. Estos estudios han dado lugar a recomendaciones como la reducción del tamaño de los grupos, la creación de comunidades de aprendizaje y sistemas de tutorías (Tinto y Pusser, 2006). Este modelo explicativo, sin embargo, también tiene varias debilidades. Metz (2004) señala que la explicación parece válida para universidades tradicionales, pero observa que podrían incidir otros factores cuando se trata de actores o ambientes disímiles, como en el caso de estudiantes que pertenecen a minorías, de estudiantes en otros rangos de edad, o de opciones universitarias atípicas, como las carreras vocacionales. En América Latina, algunos estudios han señalado que las razones para la deserción podrían ser diferentes. González Fiegehen (2007) señala como posibles factores la formación académica previa del estudiante, el nivel de aprendizaje adquirido, la falta de apoyos financieros para estudiantes e instituciones, las altas tasas de desempleo juvenil, la ausencia de una orientación vocacional, o la falta de preparación y actualización de los docentes. Los estudios latinoamericanos suelen hacer particular énfasis en las difíciles condiciones socio-económicas de la juventud. Para México, Zúñiga (2006) observa que un aspecto como el compromiso o el acoplamiento no resulta ser tan relevante como en los estudios norteamericanos. Según su estudio, resultan más importantes factores como la situación económica, la irregularidad de los estudios, o la reprobación de materias. De manera semejante, para el caso de Chile, Donoso y Schiefelbein (2007) hallaron que la situación económica del estudiante juega un papel preponderante, generando fuertes desigualdades dentro del sistema educativo.

2.4.2 ANÁLISIS DE DATOS EDUCATIVOS

El concepto de minería de datos educativos, especialmente para utilizarlos en la predicción de la deserción en las universidades, tal y como lo plantea Moulay *et al.* (2018) en la experiencia con la Universidad de Shariiaa, Fez Marruecos, dónde se busca demostrar que los académicos pueden incidir en la decisión del estudiante en no abandonar sus estudios. Se aplica el teorema de Bayes a partir de la teoría del aprendizaje estadístico para la clasificación de fortalezas y debilidades en los estudiantes quienes abandonan la universidad. Por otro lado, como puede incidir el problema de reprobación y rezago en una asignatura de primer año en el área de matemáticas, relacionado con las costumbres didácticas de los profesores y su manera de enseñanza se pueden traducir en problemas institucionales entre los cuales está la deserción de acuerdo a la experiencia de (Aparicio *et al.*, 2020) En los casos para predecir la deserción y el éxito de estudiantes, algunos estudios buscan patrones de rendimiento académico que permitieran la predicción de manera temprana de la deserción y en el caso de Dicoyskiy-Rioboo y Pedroza-Pacheco (2018) se clasifica a los estudiantes quienes llegan al final de la carrera de manera exitosa, midiendo con cuales cursos podían incidir en ese éxito o fracaso. De igual manera, se buscó una detección temprana de la deserción en el trabajo desarrollado por Rowto (2017) a través de un análisis de regresión para identificar las variables que tienen una influencia significativa en el rendimiento académico, utilizadas en un análisis discriminante lineal, en la cual se clasificaron en 3 categorías: en riesgo, aprobado o reprobado ayudando a realizar intervenciones antes de los exámenes. Por otro lado, se toma un estudio realizado por Robinson *et al.* (2016) aplicado en estudiantes de Bachillerato, donde a través del uso de modelos jerárquicos de regresión logística de efectos mixtos, se tomaron datos de impactos relativos de las características demográficas y factores académicos y de alimentación o bienestar en general para identificar que rasgos pueden ser predictores mas eficientes en cuanto a comportamiento académico y de deserción o graduación oportuna. Dando una mirada, no desde el estudiante sino desde los demás miembros de la comunidad educativa, los

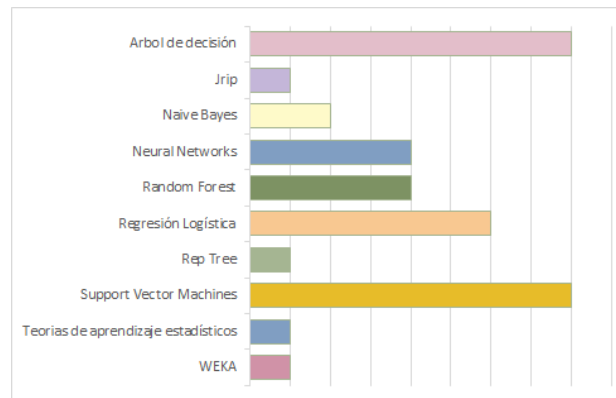
factores de deserción pueden estar relacionados con la familia, la escuela, el compromiso con el aprendizaje, el apoyo al estudiante, verificando a nivel estadístico, relaciones significativas en factores incidentes en la deserción y la percepción de los profesores y administrativos de acuerdo al estudio realizado por (Alfonso-J. *et al.*, 2019) Desde el punto de vista del factor individual o personal del individuo, se tiene una experiencia realizada por Migali y Eugenio (2017) donde desde los rasgos de personalidad de los estudiantes y la combinación con el desgaste escolar, arroja que los estudiantes con introversión son más propensos a desertar. Además de los rasgos asociados con la poca madurez psicológica del juicio, nivel bajo de responsabilidad y en general a su baja capacidad para evaluar las consecuencias a largo plazo de sus acciones y de toma de decisiones, son causales altos de la deserción. Es interesante ver como la relación entre estudiantes retenidos dentro del sistema y que abandonan con aquellos que son retenidos y no abandonan, realizado por Jimerson *et al.* (2002), está evidenciada en las características socio emocionales y conductuales tempranas que distinguen a los que abandonan, relacionando el nivel de educación materna y el rendimiento académico en la secundaria, dejando como conclusión que es muy importante el ajuste socioemocional y conductual del individuo desde edad temprana para su éxito académico a corto y largo plazo. De igual manera, se revisó la incidencia de la variable motivacional afectiva del estudiante, para predecir la autorregulación académica y su rendimiento y autoeficacia, estudio realizado por Lee *et al.* (2014), encontrando que la autorregulación académica está directamente relacionada con el interés individual para lograr el éxito académico y evitar la deserción. La minería de datos puede ayudar a la detección de comportamientos relacionados con el abandono escolar y reprobación de cursos mediante un mapeo desde el uso de base de datos relacionales, de los factores influyentes en la deserción como lo trabaja Cunha *et al.* (2016) en el Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte Brasil, dónde se propone una solución para el acceso a los datos de manera dinámica que permita a los profesores diagnosticar de manera temprana las causas de abandono y reprobación para permitir acciones de intervención para mejorar el rendimiento y reducir la deserción. Uno de los estudios más interesantes, está el de

Solis *et al.* (2018) el cual analiza el rendimiento de cuatro algoritmos de aprendizaje automático (Random Forest, Neural Networks, Support Vector Machines and Logistic Regression) con diferentes perspectivas para definir archivos de datos en la predicción de los estudiantes universitarios, se buscaba el entrenamiento del algoritmo con información de los estudiantes todos los semestres dentro de un período de tiempo determinado con una variable de clasificación que defina el estado de no abandono como graduado. Lograron una predicción correcta del 91 % de los abandonos con una sensibilidad del 87 %. Por otro lado, se tiene el estudio de Kemper *et al.* (2020) donde se realizan dos enfoques de aprendizaje automático, regresiones logísticas y árboles de decisión para predecir el abandono de estudiantes en el instituto de tecnología Karlsruhe utilizando los datos de exámenes realizados a los estudiantes para los modelos. Los árboles de decisión producen resultados ligeramente mejores que las regresiones logísticas, aunque ambos métodos producen altas precisiones de predicción de hasta un 95 % después de tres semestres. De igual manera los arboles de decisión junto con las máquinas de vectores de soporte (SVM) y redes neuronales compatibles con el kit de software WEKA se logró medir la retención de estudiantes de pregrado a través de la clasificación de datos de estudiantes, realizada por Jia y Manohar (2014), revelando que los factores más influyentes en la retención de los estudiantes estaban asociados al promedio de calificaciones y la cantidad de créditos tomadas, mejorando la clasificación de los datos a través de funciones objetivo derivadas del árbol de decisión mínimo y los algoritmos SVM para la creación de la red neuronal de dos capas y una regresión para predecir la retención. Desde el enfoque de Delen (2010) la retención afecta el ranking universitario, la reputación escolar y el bienestar financiero por lo cual se convierte en una de las prioridades más importantes para los tomadores de decisiones en las instituciones de educación superior, para lo cual es muy importante conocer las razones detrás de la deserción para predecir con precisión a los estudiantes en riesgo e intervenir de manera adecuada para retenerlos y es por eso que desde varias técnicas de minería de datos se desarrollaron modelos analíticos para predecir y explicar las razones del abandono de los estudiantes de primer año, mostrando que el modelo en conjunto funcionaba

mejor que el individual y que las variables educativas y financieras se encuentran entre los predictores más importantes del fenómeno en la población aplicada. Pal (2012) muestra que un algoritmo de aprendizaje automático puede establecer un modelo predictivo efectivo a partir de los datos de abandono escolar existentes y de esta manera identificar que estudiantes necesitan más del apoyo del programa de abandono escolar. Desde un tiempo para acá comienza a hablarse de la gestión del desempeño estudiantil, denominado según Raj y Vannan (2020), como uno de los pilares claves para las instituciones de educación superior afectando las perspectivas de carrera y clasificación del estudiante, por la cual aplica, siguiendo el camino de la analítica de aprendizaje y la minería de datos educativos mediante la aplicación de técnicas de aprendizaje automático en los datos de los estudiantes para identificar los que tenían mayor probabilidad de reprobar exámenes y por lo tanto proporcionar las intervenciones necesarias para mejorar el rendimiento de los mismos. Se usa un enfoque de minería de datos de tipo demográfico y sociales de los estudiantes, donde se comparan cinco algoritmos populares de aprendizaje automático Rep Tree, Jrip, Random Forest, Random Tree, Naive Bayes. Los resultados demostraron que el algoritmo de árbol de representación superó a otros algoritmos de aprendizaje automático al clasificar a los estudiantes que tienen más probabilidades de reprobar los exámenes. Un estudio un poco más completo acerca de la detección temprana realizado por Manrique *et al.* (2019) a través de modelos predictivos pero con una evaluación exhaustiva de la idoneidad de diferentes algoritmos y enfoques se complica por la cantidad de variables constituyen la experiencia educativa de un estudiante, sin embargo, se aplicaron en tres representaciones estudiantiles, considerando características globales basadas en características locales y series de tiempo junto con algoritmos de aprendizaje apropiados para cada una de ellas y se concluye que la deserción se predice con precisión utilizando las calificaciones de algunos cursos básicos.

Como consolidado de los diferentes artículos revisados, se obtiene información valiosa en cuanto a la tendencia de los métodos aplicados para el análisis de datos

de permanencia y deserción estudiantil así 2.4.2:



Como conclusión de la revisión realizada, se considera aplicar para el presente trabajo, la transformación de datos estudiantiles logrando una caracterización los estudiantes de tal manera que permitan dar un panorama real a través de estadísticas de los desertores del caso de estudio utilizando la metodología general de la ciencia de datos IBM y técnicas para el preprocesamiento de datos como los Sistemas de Gestión de Base de Datos, herramientas como MSEXcel, y R para el análisis exploratorio de los datos, analítica de datos, al igual que para el apoyo a la decisión multicriterio, considerando variables de tipo económico, académico e individual con la que se cuenta actualmente.

CAPÍTULO 3

DESCRIPCIÓN DETALLADA DEL PROBLEMA

En este capítulo se describe la problemática a resolver con la forma de abordar la solución, considerada mas eficiente, de acuerdo a las características propias de la misma.

3.1 PLANTEAMIENTO DEL PROBLEMA

La deserción estudiantil es un problema que afecta a todos los países del mundo, especialmente en la etapa terciaria o universitaria de acuerdo con Fontalvo (2014). El impacto social de este problema es enorme, ya que al no lograr retener a una cantidad importante de estudiantes que ingresan a las Instituciones de Educación Superior (IES) se retrasa la formación del capital humano de calidad (Centro de Estudios sobre Desarrollo Económico [CEDE], 2014), (Observatorio de Educación Superior de Medellín. [ODES], 2017). En general el problema de deserción se mide por medio de: la tasa de graduación de educación terciaria, haciendo alusión a la cobertura y deserción de los estudiantes en educación superior. La universidad se considera una institución de enseñanza superior que comprende diversas facultades, colegios, institutos, departamentos, centros de investigación, escuelas profesionales,

etcétera, y que otorga los grados académicos correspondientes (Real Academia Española (RAE) , 2001). Dichas instituciones se ven en la tarea de definir sus procesos estratégicos, operativos y de apoyo, desde una perspectiva de planeación, ejecución, verificación, y toma de acciones que propendan por generar estados educativos de alta calidad y su objetivo fundamental es la formación de profesionistas y científicos de alta competitividad. De acuerdo al gráfico 3.1 siguiente, el promedio de jóvenes que termina la universidad en los países de la OCDE es inferior al 40 por ciento, dos factores que inciden en este problema son la baja cobertura y la deserción escolar.

Gráfico Nro 1: Población de estudiantes quienes terminan la Universidad Países de la OECD

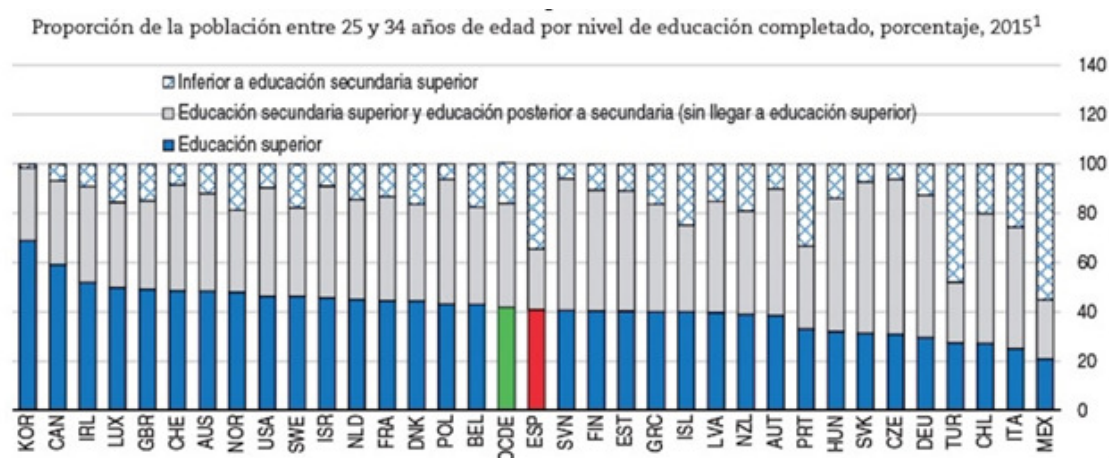


FIGURA 3.1: Poblacion de graduados OCDE 2015 2018.

Desde un Comparativo tasas de graduación de la OCDE, 2015, 2016, 2017, 2018 se tiene que el porcentaje de graduación oportuna pasó de 41 por ciento a 75 por ciento entre el 2013 al 2016, atribuyendo que el porcentaje restante responde a desertores.

En la siguiente tabla 3.1 se muestran datos consolidados de deserción estudiantil en países de la OCDE

TABLA 3.1: Índice de abandono de estudiantes en educación terciaria a 2017.

País de la OECD	Índice de abandono
Belgium	22 %
Slovenia	19 %
Iceland	18 %
Lithuania	18 %
Chile	17 %
Sweedden	15 %
Austria	13 %
Australia	12 %
Norway	12 %
Portugal	12 %
Netherlands	12 %
Estonia	11 %
Brasil	11 %
New Zeland	10 %
COLOMBIA	10 %
France	8 %
Switzerland	8 %
Israel	8 %
United Kingdom	8 %
Finland	8 %
USA	6 %

3.2 CARACTERÍSTICAS DEL PROBLEMA

La deserción en Colombia, se aborda en la diferentes IES, considerando los factores definidos en la Ministerio de Educación Nacional Colombiano (2015) los cuales se visualizan en la siguiente Figura 3.2: Factores incidentes en la deserción

estudiantil



FIGURA 3.2: Factores Incidentes. Ministerio de Educación Nacional Colombiano

De acuerdo a la literatura consultada, el impacto de los factores y la manera de abordarlos, se complica al intentar tomar todos en cuenta de manera simultánea por su naturaleza diversa, la dificultad de establecer las medidas de dicho impacto por su naturaleza subjetiva, haciendo mas complejo la toma de decisiones, ya que se requiere información sobre los estudiantes, que en muchos casos se llevan para estadística pero no para desarrollar estrategias de acompañamiento o mitigación del abandono o detección temprana de posibles desertores. Para ello, se requiere desarrollar sistemas de apoyo a la decisión efectivos y eficaces que permitan intervenir a tiempo en estudiantes desertores potenciales. Cabe resaltar otro aspecto importante incidente relacionado con la capacitación del personal docente para detectar y apoyar

a posibles desertores. Según Carvajal-Olaya y Trejos-Carpintero (2016) la deserción obedece a un conjunto de factores tales como: individuales (antropológicos), socio-económicos, académicos e institucionales, los cuales el factor de mayor incidencia en la deserción es la mortalidad académica, cuya participación es en promedio del 60 % sobre la deserción total. Por último, señalan que deficiencias en los sistemas de información dificultan el seguimiento a las estrategias sobre deserción. La relación de los servicios académicos ofrecidos a los estudiantes desde la relación de profesores y el proceso de caracterización estudiantil en cuanto a estilos de aprendizaje y variables de tipo académicas y antropológica, deben ser de total interés dentro de la generación de planes de estudio y oferta académica en general en las Instituciones de Educación Superior, además de propiciar un clima de comunicación abierta y sincera para los estudiantes, y lograr una cultura de seguimiento y acompañamiento transparente logrando que los estudiantes expresen con libertad sus necesidades para lograr una permanencia con calidad logrando la graduación oportuna de los mismos. En este sentido, sería interesante enfocar la institución como una cadena de suministro donde la más importante fuente de valor son precisamente sus estudiantes por lo que toda su gestión debe centrarse en prestarles una atención de excelencia y mantenerlos en la institución para que terminen sus estudios con calidad. Esto implicaría a la institución orientar parte de sus procesos y procedimientos a la gestión de la permanencia estudiantil con calidad.

La autora considera que al generar estrategias de mitigación de la deserción escolar se debe dar respuesta a las siguientes preguntas:

- ¿Por qué desertan los estudiantes universitarios?
- ¿Cómo se aborda la deserción estudiantil desde las instituciones universitarias?

CAPÍTULO 4

METODOLOGÍA PROPUESTA

En este capítulo se describe de forma detallada la metodología propuesta y cada una de sus fases.

Para el desarrollo de la misma se decide trabajar en dos etapas. La primera etapa se desarrolla todo lo relacionado con un Diagnóstico y en la segunda etapa se aborda lo relacionado con la Intervención desde los resultados.

El objetivo de la etapa de Intervención es apoyar la toma de decisiones para diseñar un modelo de acompañamiento y seguimiento estudiantil óptimo de acuerdo la situación particular del centro en el cual se realiza la intervención.

4.1 MODELO PROPUESTO PARA GESTIÓN DE LA PERMANENCIA DE CALIDAD EN UNA INSTITUCIÓN DE EDUCACIÓN SUPERIOR

El Ministerio de Educación Nacional Colombiano (2015) en su **Guía para la implementación del modelo de gestión para la permanencia y graduación estudiantil en instituciones de educación superior**, al tener en cuenta la complejidad y la diversidad de factores que afectan las posibilidades para que

un estudiante pueda permanecer en la educación superior, desarrolló, de manera progresiva, y con la participación de diferentes actores del sector educativo, un modelo que aborda el tema desde las diferentes dimensiones, en el que sincroniza las mayores oportunidades de acceso con la generación de condiciones que garanticen el logro académico final del estudiante. El modelo presentado fue resultado de un trabajo continuo y un riguroso estudio a partir de la medición del fenómeno que se ha realizado desde el Sistema para la Prevención de la Deserción de la educación superior (SPADIES) y que se ha combinado con las estrategias que las Instituciones de Educación Superior han venido adelantando, lo cual logró una comprensión cada vez más integral de la problemática, y su objetivo consistió en definir lineamientos, estrategias e instrumentos en ocho ejes de trabajo, para fortalecer la capacidad institucional en el diseño, ejecución y evaluación de políticas y programas de fomento de la permanencia y la graduación estudiantil de manera integral. En el gráfico 4.1 se presentan los ocho ejes del modelo:



FIGURA 4.1: Modelo propuesto por el Ministerio de Educación Nacional Colombiano 2012

Cada eje describe a continuación su objetivo de acuerdo con lo expresado por el Ministerio de Educación Nacional Colombiano (2015):

- Posicionamiento y formalización: Generar, en los diferentes miembros de la comunidad académica, como directivos, docentes, administrativos, estudiantes y familias, la cultura de la promoción en torno a la permanencia y graduación estudiantil, a partir de la definición y aprobación de **políticas, estrategias, procesos y procedimientos** que involucren, guíen y articulen la actuación de los miembros y soporten las estrategias de apoyo a los estudiantes.
- Cultura de la información: Hacer un seguimiento periódico a la deserción estudiantil, desde la detección de alertas tempranas para identificar el riesgo de abandono por estudiante, estimar los costos que este genera y evaluar el impacto de las estrategias y programas que desarrollan las instituciones. Se busca que las IES cuenten con información completa, veraz y oportuna, que les permita conocer su situación institucional y diseñar estrategias específicas y pertinentes a su contexto.
- Mejoramiento de la calidad: Integrar procesos de investigación académica y evaluación de docentes para desarrollar competencias pedagógicas y actualizar metodologías, que faciliten el aprendizaje de los estudiantes y optimicen el desempeño y la excelencia académica, con la implementación y uso de tecnologías de la información y de las comunicaciones. Igualmente, visibilizar la permanencia y la graduación en los procesos de registro calificado y acreditación de calidad.
- Trabajo conjunto con las Instituciones de educación media (IEM): Crear alianzas que fortalezcan el trabajo conjunto entre las Instituciones de Educación Media, las secretarías de educación y las IES para:
 - Movilizar la demanda a la educación superior de los estudiantes de media con estrategias de nivelación académica para el mejoramiento de las competencias,

- Fomentar ejercicios de reflexión como estrategias de orientación socioocupacional en los componentes del autoconocimiento, oferta de la educación superior y el conocimiento del mundo del trabajo, en los estudiantes que se encuentran en la educación media y
 - Trabajo articulado para la cualificación de los docentes de la Educación Media.
-
- Programa de apoyo a estudiantes: Definir y articular programas de acompañamiento integral a estudiantes, con el fin de aumentar su cobertura e impacto. Se busca generar estrategias que le permitan a la institución orientar al estudiante, desde el inicio de su programa académico hasta la culminación exitosa del mismo. En ella se incluye la evaluación permanente de los programas en marcha, así como una revisión sobre su adaptación permanente a las cambiantes condiciones estudiantiles.
 - Compromiso con el núcleo familiar: Promover la participación y el aporte del núcleo familiar en el acompañamiento, la motivación y la construcción de proyecto académico del estudiante, para la culminación efectiva de su plan de estudios, a partir de la sensibilización a la familia frente a la función que ejerce como parte fundamental de la red de apoyo psicosocial del estudiante.
 - Gestión de recursos: Fomentar la participación y el aporte de diversos agentes, para el apoyo a estudiantes identificados en riesgo por factor financiero. Se orienta a la generación de acciones desde dos ejes de trabajo:
 1. Asignación de recursos propios de la institución hacia los programas de apoyo a la permanencia para su continuidad y sostenibilidad y
 2. La consecución de recursos mediante la vinculación del sector público o privado, para el patrocinio a estudiantes en condiciones de vulnerabilidad.
 - Trabajo colaborativo: Aunar esfuerzos entre las IES, para definir, ejecutar y evaluar estrategias que permitan disminuir la tasa de deserción estudiantil y fortalecer la capacidad institucional para el fomento de la permanencia y

graduación estudiantil, mediante la transferencia de conocimiento, experiencias exitosas, metodologías y recursos.

En cada uno de los ejes se definen un conjunto de herramientas desde las experiencias de las universidades colombianas, sin embargo, en el marco del proyecto se realiza un análisis por parte de la doctorante, donde se sugieren ciertas actividades y alternativas que se pueden aplicar para cada uno de ellos así:

1. Sensibilización - Socialización 4.2:



FIGURA 4.2: Eje de Sensibilización

2. Cultura de la Información 4.3: Desarrollar, actualizar y afinar desde el sistema de información institucional:



FIGURA 4.3: Eje Cultura de Información.

3. Mejoramiento de la calidad académica:

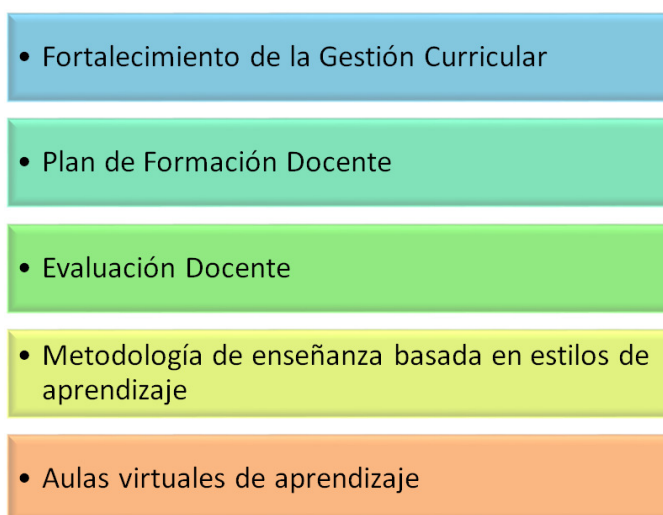


FIGURA 4.4: Eje Mejoramiento Calidad Académica P1

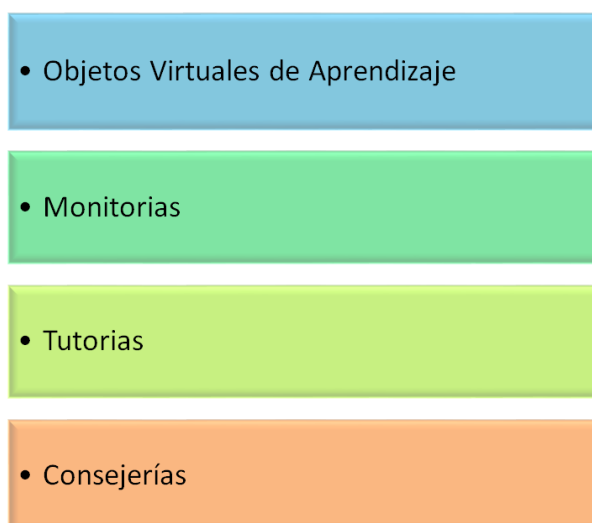


FIGURA 4.5: Eje Mejoramiento Calidad Académica P2

4. Trabajo en grupo con las Instituciones de Educación Media (IEM) 4.6



FIGURA 4.6: Eje Trabajo en grupo con Instituciones de Educación Media

5. Programa de apoyo para los estudiantes 4.7



FIGURA 4.7: Eje Programa de apoyo a estudiantes

6. Compromiso con el núcleo familiar



FIGURA 4.8: Eje Compromiso con el núcleo familiar

7. Gestión de recursos:



FIGURA 4.9: Eje Gestión de Recursos

8. Trabajo colaborativo entre Instituciones de Educación Superior:



FIGURA 4.10: Eje Trabajo Colaborativo con Instituciones de educación superior

Considerando lo planteado en cada eje, las universidades pueden crear una batería de indicadores en cada uno de ellos, e iniciar un proceso de implementación acorde con el tipo de población estudiantil con la que cuentan, debido que es necesario e imperante, conocer las características de los estudiantes para poder generar las estrategias y acciones efectivas para el grupo poblacional.

4.2 ENFOQUE METODOLÓGICO

Según la literatura revisada el problema que se aborda es complejo por la cantidad y diversidad de información que se requiere procesar para poder aportar soluciones eficaces y por el tipo de decisiones que se deben tomar, como por ejemplo el diseño de estrategias y políticas para mitigar la deserción, que son fundamentalmente de tipo no estructurado (que resultan los problemas más difíciles en la toma de decisiones). Estas dos características son las que motivan la formulación de un enfoque metodológico novedoso en el que se integran el **”método científico, la ciencia de datos y el apoyo a la decisión multicriterio”**. Esta integra la Metodología General de Apoyo a la Decisión Multicriterio (*MGADM*) que fue propuesta por el investigador Simon (1960) y la metodología fundamental para la ciencia de datos de IBM (*MGCD-IBM*) propuesta por Rollins (2015). De igual manera, se consideran aspectos relevantes obtenidos a partir de la revisión de la literatura realizada. Ambas descritas en lo fundamental en el capítulo II de este documento.

1. **El Metodo científico:** Garantiza la veracidad, confiabilidad, reproducibilidad, extensibilidad o generalización de los resultados.
2. **La ciencia de datos:** Facilita la eficacia y eficiencia en la captura, validación, estandarización, modelamiento, análisis e implementación de resultados para la toma de decisiones basada en datos.
3. **Metodología General para el apoyo a la decisión multicriterio:** Al ser un problema complejo donde intervienen múltiples factores, características y

dimensiones, con elementos subjetivos que son fundamentales para resolver el problema, esta es la única metodología que garantiza la eficiencia, eficacia y racionalidad en el proceso de toma de decisiones.

4.2.1 REQUERIMIENTOS PARA LA INTERVENCIÓN

La aplicación de la metodología en su fase de intervención en una organización requiere que esta cumpla con una serie de requerimientos mínimos previos a la intervención:

- Existencia de un registro automatizado con seguimiento a los estudiantes en cuanto a:
 - (i) situación personal (si trabaja, tiene dependientes económicos, es dependiente económico, clase social a la que pertenece, etc.) ,
 - (ii) situación académica (ingresos, abandonos, graduación, calificaciones),
 - (iii) situación económica (si el estudiante recibe algún tipo de apoyo, y periodos o semestres y años en los que ha recibido estos apoyos). Lo mínimo que se requiere es que se lleve un registro de las deserciones o abandonos de los estudiantes por periodo académico (semestre, etc.), así como los datos su trayectoria académica (registros por periodo, calificaciones).
- Estrategias de tutoría y políticas implementadas por la institución para mitigar el abandono, así como un registro electrónico de esa actividad (tipo de acompañamiento recibido, resultados, periodo o semestre y año del acompañamiento). Este requisito no es obligatorio.

4.3 ETAPA DE DIAGNÓSTICO

El objetivo de la etapa de Diagnóstico es la identificación del nivel actual de madurez en la gestión y solución al problema de la deserción estudiantil de la institución que se intervendrá aplicando la metodología propuesta en este trabajo, para:

1. Certificar un nivel de madurez (eficacia) en la gestión y mitigación de la deserción escolar.
2. Establecer si cumple o no los requisitos mínimos par aplicar la metodología en su etapa de intervención.
3. Emitir recomendaciones para resolver las áreas de oportunidad detectadas en la mitigación de la deserción escolar.

En la subsección siguiente se define el problema de la evaluación diagnóstico como un problema de evaluación multicriterio.

4.3.1 DIAGNÓSTICO COMO PROBLEMA DE DECISIÓN MULTICRITERIO

En <https://www.dictionary.com/browse/diagnosis> se define "Diagnóstico" como el proceso de examinar y analizar las causas de una situación problemática y las recomendaciones o soluciones que se derivan de este proceso.

En el marco de esta investigación se define el problema del diagnóstico de la gestión para la mitigación de la deserción estudiantil en un centro de educación superior como un problema de evaluación multicriterio

Zopounidis y Doumpos (2002). Esto en base a los resultados del diagnóstico

expuestos arriba 1-3. La elección de este tipo de problema de decisión multicriterio se justifica con el hecho de que las alternativas no están disponibles a priori y también porque de manera natural lo que se persigue es asignar cada alternativa a una clase predefinida de evaluación. Estas características favorecen la elección del tipo de problema de evaluación contra los problemas de jerarquización o selección Roy.

El elemento central en el problema es el tomador de decisiones, que en este caso es la entidad evaluadora (si es un auto diagnóstico entonces es la propia institución que es evaluada representada por quien o quienes dirigen la comisión o comité de autoevaluación). Es importante reconocer al tomador de decisiones ya que representa al responsable directo de las consecuencias de la toma de decisiones (puede ser un individuo o un grupo de individuos). Es importante identificar correctamente al tomador de decisiones, pues son sus preferencias las que deben modelarse e ingresar en el modelo.

Para iniciar la descripción del resto de los elementos del problema se definen los niveles de gestión de interés que se establecerán al final del diagnóstico:

1. *Reconocimiento*: En este nivel la institución ha determinado que efectivamente la deserción estudiantil constituye un problema que afecta sus intereses estratégicos, sin embargo aún no concreta estrategias ni mecanismos para mitigar su alcance y efecto, o lo implementado no tiene un impacto positivo significativo en la mitigación de la decisión.
2. *En desarrollo*: En este nivel la institución ya ha diseñado y/o implementado estrategias de mitigación, aunque aún estas están en un proceso de perfeccionamiento, consolidación.
3. *Madurez*: En este nivel la institución ha desarrollado, implementado y consolidado estrategias de mitigación, cuenta con mecanismos, tiempo y datos suficientes para evaluar el impacto de estas estrategias. Aunque el proceso está en un continuo perfeccionamiento.

Existe una relación de preferencia natural entre estos niveles, que se expresa como:

Reconocimiento \prec En desarrollo \prec Madurez.

Lo que significa que el nivel de Reconocimiento es menos apreciado, mientras que el nivel de Madurez es el Mas apreciado.

En Vicerrectoría (2018) académica de la Universidad Simón Bolívar se propuso un modelo de permanencia institucional, sobre este y lo encontrado en la literatura revisada se propone la descomposición de objetivos para establecer el conjunto de criterios según la metodología propuesta por Bell *et al.* (1977).

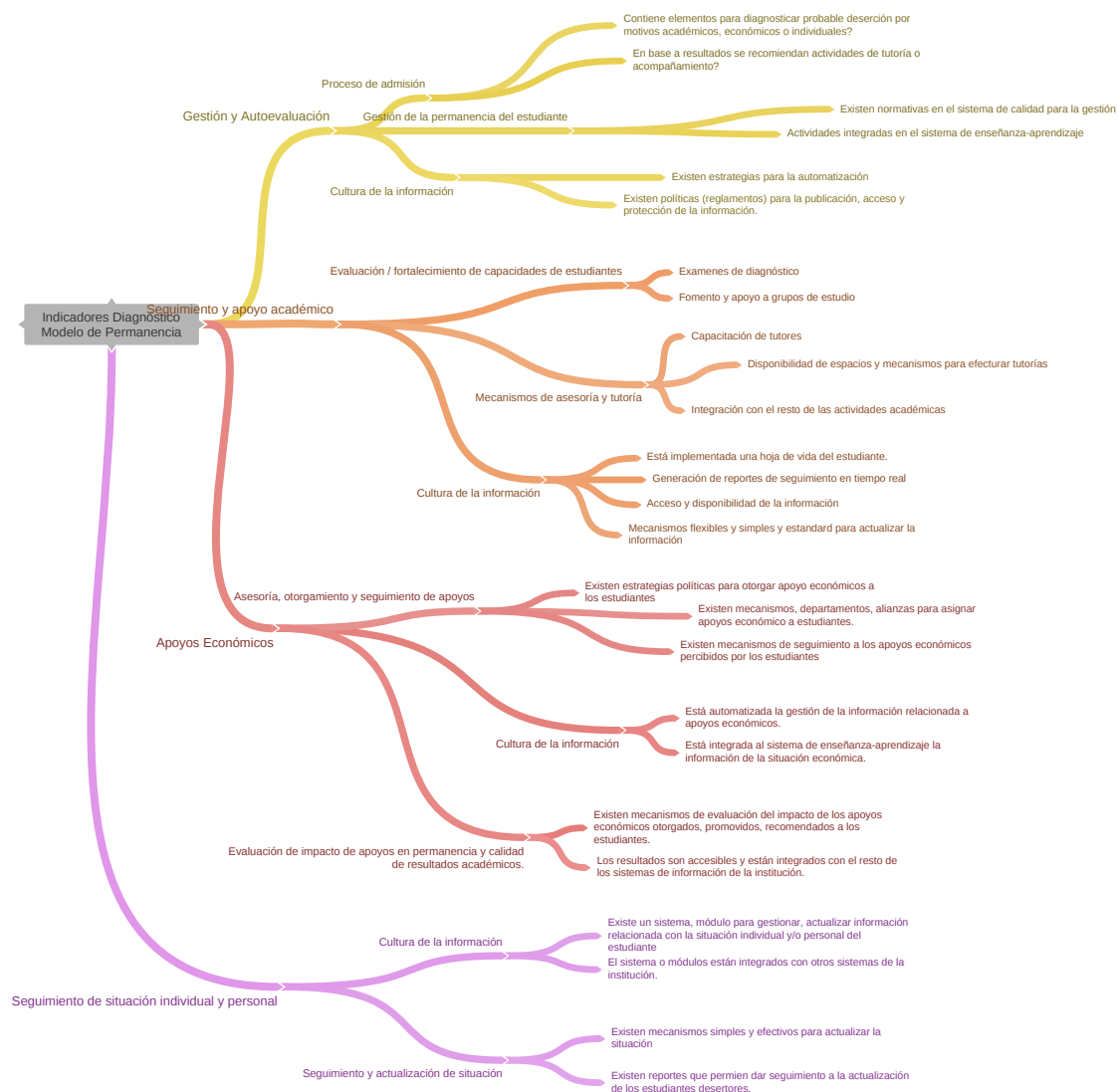


FIGURA 4.11: Jerarquía de Criterios

En base a la jerarquía mostrada en la figura 4.11 se proponen los siguientes criterios compuestos:

■ Gestión y Autoevaluación.

1. Proceso de admisión.

- Identificación de posibles desertores.
- Recomendación de acompañamiento para posibles desertores.

2. Gestión de la permanencia.

- Integrada en el sistema de calidad.
- Integrada en el sistema de enseñanza-aprendizaje.

- **Seguimiento y apoyo académico.**

1. Evaluación o fortalecimiento de capacidades de estudiantes.

- Se aplican exámenes de diagnóstico.
- Fomento, apoyo a grupos de estudio.

2. Mecanismos para asesorías y tutorías.

- Capacitación de tutores.
- Disponibilidad de espacios y recursos para tutorías o asesorías.
- Integración con otras actividades académicas.

- **Apoyos económicos.**

1. Asesoría, otorgamiento y seguimiento de apoyos.

- Pertinencia de estrategias, políticas para otorgar apoyos.
- Mecanismos para la facilitar la asignación de apoyos externos a estudiantes.
- Mecanismos de seguimiento a los apoyos asignados a los estudiantes.

2. Pertinencia de evaluación de impacto de apoyos en permanencia y calidad de resultados académicos.

- Pertinencia de Evaluación del impacto en la permanencia de los apoyos económicos otorgados.
- Los resultados de la evaluación del impacto son accesibles y están integrados con el sistema de enseñanza-aprendizaje.

- **Seguimiento de situación individual o personal del estudiante.**

1. Pertinencia de mecanismos simples y efectivos para actualizar la situación individual o personal del estudiante cuando se ocupe.

2. Pertinencia de mecanismos para dar seguimiento a la situación individual o personal del estudiante cuando se ocupe.

■ **Cultura de la información.**

1. En la gestión y autoevaluación.
 - a) Pertinencia de estrategias para la automatización de la gestión de la información relacionada con la gestión de la Permanencia.
 - b) Pertinencia de políticas (reglamentos) para la publicación, acceso y protección de la información.
2. En el seguimiento y apoyo académico.
 - a) Está implementada la hoja de vida del estudiante.
 - b) Generación de reportes de seguimiento en tiempo real.
 - c) Acceso y disponibilidad a la información.
 - d) Mecanismos simples y flexibles para actualizar la información.
3. En apoyos económicos.
 - a) Automatización de la gestión de la información relacionada con apoyos económicos.
 - b) Integración al sistema de enseñanza-aprendizaje la información sobre apoyos económicos.
4. En seguimiento de la situación individual o personal.
 - a) Existe un sistema, módulo para gestionar, actualizar información relacionada con la situación individual y/o personal del estudiante.
 - b) Integración de la gestión de la información del seguimiento con otros sistemas informáticos de la institución.

El paso siguiente consiste en establecer escalas de preferencias para los criterios o atributos de decisión, estas escalas deben tener un significado claro para el tomador de decisión, y ser simples para facilitar la interpretación de la evaluación. En este

caso se ha escogido una escala de tipo nominal que solo aporta información ordinal, este tipo de escala son las más sencillas y sólo aportan información de orden entre las alternativas. Si bien este tipo de escalas son limitadas en cuanto a las operaciones que se realizan sobre ellas, en realidad no se requiere más, puesto que el modelo de evaluación multicriterio no precisa realizar operaciones o distinguir intensidad de preferencias en la comparación entre alternativas.

Se propone utilizar una escala estandar con cuatro niveles de preferencia:

- *Sin impacto*: Es el nivel más bajo de preferencia que indica ausencia de una característica correspondiente a un criterio o atributo, o su nivel de implementación aún no genera impacto observable en el sistema.
- *Impacto mínimo*: Este nivel indica que una característica asociada a un criterio o atributo si está presente pero que aun tiene muchas áreas de oportunidad que deben abordarse para que tenga un real impacto positivo en el sistema.
- *Impacto suficiente*: Indica que la característica tiene un nivel de implementación que tiene un impacto positivo en el sistema aunque aun tiene áreas de oportunidad.
- *Impacto importante*: Este nivel indica que la característica tiene un impacto positivo importante en el sistema, y que aunque tiene áreas de oportunidad en su implementación es una característica madura.

Las alternativas del problema son un elemento interesante de este problema, ya que no existen de antemano, sino que son el resultado del proceso de evaluación en la etapa de diagnóstico. Y dada la cantidad de resultados posibles la cantidad de alternativas es muy grande (resulta la combinación de los niveles de la escala de preferencias en los 22 atributos establecidos (ver 4.11) o lo que es lo mismo 22^4 , lo que dificulta o hace casi imposible la asignación directa de las consecuencias a cada alternativa.

En los problemas de evaluación multicriterio existe una forma de circunvenir la dificultad comentada en el párrafo anterior y esta consiste en definir perfiles o .alternativas de referencia” (ficticias o reales) que establecen las fronteras entre las categorías de evaluación. La definición de estos perfiles es un paso decisión en la modelación ya que inciden en los resultados finales de manera importante. Para casos complejos es posible definir múltiples perfiles para cada categoría de evaluación, este es un problema complejo por la cantidad de atributos así como por la interpretación de las categorías por ello se decidió definir múltiples perfiles para definir las categorías.

Cuando se emplean múltiples perfiles para definir una categoría no debe ser posible establecer una preferencia débil o parcial entre ellos, pero si pueden ser incomparables o indiferentes.

Para definir los perfiles se deben cumplir dos requisitos ?:

- Toda alternativa debe quedar asignada a una única categoría, no pueden quedar alternativas sin clasificar.
- Las alternativas de referencia que definen el perfil de una categoría a deben ser preferidas a aquellas alternativas de referencia que definen los perfiles de las categorías b que tienen un nivel más bajo de preferencia que a , en otras palabras para las cuales $b \prec a$.

Una vez definidos los perfiles se asigna una alternativa a una categoría si esta es indiferente o preferida a al menos un perfil y no existe un perfil que define la categoría que sea preferido a la alternativa en cuestión. Las definiciones entre dos alternativas de: preferencia, indiferencia o incomparabilidad dependerán del método de decisión empleado. En esta investigación se propone emplear el método ELECTRE TRI ? que es los más utilizados por estar basado en principio simples que pueden ser perfectamente comprendidos por los tomadores de decisión.

Definición de la categoría *Reconocimiento*: Este es el nivel de madurez más bajo en la evaluación del sistema, se identifica porque no existen estrategias, tácticas

o mecanismos implementados para mitigar la deserción o de haberlos no tienen un impacto positivo observable y documentado. Este nivel queda definido por los perfiles que definen a la categoría *En desarrollo*.

Definición de la categoría *En desarrollo* Este nivel se caracteriza fundamentalmente porque existen evidencias de que hay Estrategias, políticas o mecanismos implementados y su impacto positivo para mitigar la deserción estudiantil.

Al definir el perfil hay que establecer cuales son los niveles mínimos que se requieren para esta categoría y que la diferencian de la categoría *Reconocimiento*, en este sentido lo que marca la diferencia es el impacto de la estrategias, políticas o mecanismos implementados reflejado en las evidencias disponibles.

Hay varias formas de establecer el impacto pero todas deben evidenciarse con registros que muestren la disminución de la deserción en varios periodos, al menos dos periodos consecutivos.

En relación al **objetivo a)** de esta fase se definen el o los perfiles que definen esta categoría, para los cuales se han escogido dos indicadores que deben mostrar un cumplimiento mínimo:

- Dentro de **Gestión y Autoevaluación** se considera: *Gestión de la permanencia*.
 - Integrada en el sistema de enseñanza-aprendizaje, se exige un nivel de *Impacto suficiente*.
- Dentro de **Seguimiento y apoyo académico** se considera *Mecanismos para asesorías y tutorías*.
 - Disponibilidad de espacios y recursos para tutorías o asesorías, se exige un nivel de *Impacto suficiente*.
 - Integración con otras actividades académicas, se exige un nivel de *Impacto suficiente*.

Se define otro perfil (no obligatorio para la categoría) para establecer si la institución cumple con los requerimientos mínimos para aplicar la metodología en su fase de intervención **objetivo b)** descrito arriba), este perfil se relaciona directamente con el atributo de **Cultura de la información** en el subatributo *En el seguimiento y apoyo académico.*:

1. Está implementada la hoja de vida del estudiante, se exige un nivel de *Impacto mínimo*.
2. Generación de reportes de seguimiento en tiempo real, se exige un nivel de *Impacto mínimo*.
3. Acceso y disponibilidad a la información, se exige un nivel de *Impacto suficiente*.
4. Mecanismos simples y flexibles para actualizar la información, se exige un nivel de *Impacto suficiente*.

Definición de la categoría *Madurez* Este es el nivel más alto de desarrollo de una institución, para ser catalogada en un nivel de **Madurez** debe evidenciar resultados importantes y sostenidos en la disminución de los indicadores importantes de la deserción estudiantil, que parten desde una gestión comprometida por parte de la administración y dirección hasta una cultura de información madura con una alta integración con otros sistemas del proceso de enseñanza-aprendizaje.

Para definir este nivel se consideran:

- **Gestión y Autoevaluación.**: todos los subcriterios deben tener una evaluación de al menos *Impacto suficiente*.
- **Seguimiento y apoyo académico**: todos los subcriterios deben tener una evaluación de *Impacto Importante*.

- **Apoyos económicos.:** todos los subcriterios deben tener una evaluación de al menos *Impacto suficiente*.
- **Seguimiento de situación individual o personal del estudiante.:** todos los subcriterios deben tener una evaluación de al menos *Impacto suficiente*.
- **Cultura de la información:** todos los subcriterios deben tener una evaluación de *Impacto Importante*.

Para efectuar la evaluación se recomienda primero recolectar la evidencia requerida por cada indicador, organizarla para luego llevar a cabo la evaluación, dar cumplimiento a los objetivos a) y b), la asignación de la categoría se inicia confrontando el resultado de la evaluación respecto a la categoría mas preferida hasta la menos preferida. Se le asignará a la evaluación la primera categoría que acredite la pertenencia a la misma.

Una vez realizada la asignación y estén cumplidos los objetivos a) y b) se procede a redactar las recomendaciones y dar cumplimiento al objetivo c).

En el **Anexo 1** se incluye una plantilla propuesta en MS Excel para registrar la información recabada en esta etapa de diagnóstico, generar la evaluación así como registrar las observaciones y recomendaciones.

4.4 ETAPA DE INTERVENCIÓN

Como se ha mencionado anteriormente el propósito de esta etapa es apoyar la decisión en la formulación de estrategias, tácticas, mecanismos y actividades para mitigar la deserción escolar en un centro de educación superior. Esto se lleva a cabo desde la perspectiva de la decisión multicriterio y la ciencia de datos para que la decisión sea coherente con la evidencia aportada por la información (decisión soportada por datos, o Data Driven Decision como se ha acuñado el término en inglés) y consecuente con los objetivos, misión y visión de la institución.

Esta característica hace que la intervención ofrezca como resultado evidencias mas que una recomendación concreta para apoyar la toma de decisiones, ya que el complejo problema que se aborda tiene muchas realizaciones que dependen de la situación específica de la institución intervenida (cuyo estado se actual se establece en la fase de diagnóstico de la metodología propuesta). También ofrece una vía para organizar la recolección de evidencias y definir métricas pertinentes para evaluar el impacto de las estrategias, tácticas y mecanismos implementados para instituir un plan de mejora continua y avanzar así hacia una cultura de la excelencia en sus actividades.

En las subsecciones siguientes se definen y detallan los pasos recomendados para su aplicación en esta etapa.

4.4.1 DEFINICIÓN Y ESTRUCTURA DEL PROBLEMA DE DECISIÓN. PASO 1

Este paso es una combinación de las fases o pasos de las metodologías:

MGADM (Metodología General de Apoyo a la Decisión Multicriterio): Se aplica para la definición del problema de decisión: Inteligencia. **MGCD-IBM** (Metodología General de la Ciencia de Datos propuesta por IBM): Se aplica para la definición de los elementos del problema de decisión. (Entendimiento del negocio, Aproximación analítica, Requerimientos de datos)

En este paso se establecen los elementos principales que constituyen el problema de decisión en cuestión, los modelos requeridos para el análisis de los datos, síntesis de la información para facilitar la aplicación de métodos de apoyo a la decisión multicriterio, requerimientos de datos, así como las fuentes desde donde se obtendrán los datos requeridos del problema.

4.4.1.1 DEFINICIÓN DEL PROBLEMA DE DECISIÓN. (PASO 1.1)

Lo primero que debe hacerse es definir claramente el problema de decisión:

La definición del problema de decisión concreto esta fuertemente relacionada con la experiencia acumulada en la mitigación de la deserción en otras instituciones, y con los resultados y las recomendaciones que resulten de la etapa de diagnóstico de la metodología. Estos dos aspectos establecen dos componentes fundamentales del problema de decisión:

1. General: esta componente es independiente de la institución y su situación respecto a la deserción al momento de la intervención, se relacionará directamente con la deserción estudiantil, así como las estrategias, políticas, mecanismos y actividades típicas que se llevan a cabo para mitigarla. Las mismas aparecen ampliamente reportadas en la literatura científica publicada, reportes institucionales, lineamientos de organizaciones reguladoras de la educación, etc. Esta componente aporta la estructura del problema (a un nivel abstracto).
2. Específica: esta componente es altamente dependiente de la institución intervenida, así como de los resultados de la aplicación de la fase de diagnóstico. Esta componente aporta los elementos de la estructura (la instancia específica del problema).

Considerando estas dos componentes el problema de decisión que se aborda puede definirse como:

- Problema general: Caracterizar a un estudiante desertor atendiendo a factores académicos, económicos, institucionales, personales.
- Problema específico: Establecer un proceso robusto, flexible, eficiente para apoyar la toma de decisiones en la mitigación de la deserción estudiantil desarrollando *estrategias, tácticas, mecanismos y actividades* que estén totalmente

alineados con la visión estratégica de la institución, su visión a futuro y con las recomendaciones pertinentes resultantes del proceso de diagnóstico (Etapa inicial de la metodología propuesta), planeando el mismo en función de los recursos puedan destinarse a este propósito en el marco de tiempo que se lleve a cabo dicho proceso.

Esta metodología se enfoca en el problema general descrito anteriormente.

4.4.1.2 ESTRUCTURA DEL PROBLEMA DE DECISIÓN. (PASO 1.2)

El problema de decisión descrito en la sub sección anterior se corresponde con el tipo de problema de clasificación multicriterio”, descrito en el capítulo 2. Los elementos de este tipo de problema son muy similares a los elementos del problema de evaluación multicriterio descrito en la fase de diagnóstico, sin embargo, la diferencia entre ambos problemas radica en que en este problema no existe una relación de preferencia entre las categorías. No obstante es posible definir esta relación para separar ambas categorías si se asume que el propósito principal en la solución de este problema es caracterizar a los desertores. En lo que sigue, se definirán los elementos del problema de decisión.

Categorías En este problema son dos las categorías (aunque pueden definirse mas en función de los intereses de la institución): (a) DESERTOR, designa a un estudiante que es considerado como desertor por la institución en un periodo de tiempo, no se matriculó en ese periodo (la institución define el periodo mínimo establecido para considerar la deserción, por ejemplo en Colombia se considera desertor a un estudiante que no se matricula por dos periodos consecutivos), (b) NO DESERTOR: se designa a un estudiante que aparece matriculado en un periodo concreto.

Tomador de decisión La institución, representada por la dirección, junta de gobierno, u otro órgano designado en lo adelante se le denominará simplemente DM

(Decisor Maker en inglés, como usualmente se le denomina en el argot de MCDM). Como se ve el DM es un rol de un individuo, o grupo de individuos, que representa los intereses de la institución de forma directa y que responde por las consecuencias de la toma de decisiones.

Atributos o criterios Se derivan del modelo de permanencia propuesto, en concreto de los ejes: académico, económico, institucional y personal. Así como de los indicadores propuestos en el mencionado modelo de permanencia. La definición de los atributos concretos corre a cuenta del DM. Algunos atributos relevantes, por ejemplo, pueden ser, por cada periodo del registro del ciclo de vida del estudiante en la institución:

- **Rendimiento académico o Calificaciones:** valoración cualitativa del rendimiento académicos del estudiante. Al que se le puede asignar la escala *No matriculado, Reprobado, Aceptable, Bueno, Excelente*.
- **Acompañamiento:** si recibió algún tipo de acompañamiento académico durante el periodo (asesoría, tutoría, asistencia psicológica, etc.). La escala natural es *Si, No*
- **Apoyo económico:** si ha recibido apoyo económico en el periodo o no, de manera natural se le asigna la escala cualitativa *Si, No*.
- **Estrato:** Estrato social al que pertenece, por ejemplo en Colombia es una escala del *1 al 6*, otra escala mas general es: *Pobre, Clase media, Clase media alta, Clase alta*. Igualmente pueden definirse otras escalas de preferencia.
- **Situación personal:** Si es casado, Si tiene hijos, Si trabaja, etc. El propósito de este atributo es establecer si su situación personal puede interferir con su permanencia en la institución. Por ejemplo, podría definirse la escala *Muy poco, Algo, Mucho*.
- **Facultad o carrera que cursa:** este es un atributo al cual es muy difícil asignarle una escala de preferencia en general, pero si puede determinarse su

incidencia en los porcentajes de deserción en una institución en particular.

La definición final de los atributos y sus escalas (conjuntos de valores relevantes, con una relación de preferencia definida o no) se realiza en el paso siguiente de la metodología (Paso 2), una vez que hayan obtenido los datos relevantes y se realice el análisis exploratorio de los mismos.

Alternativas Son los registros existentes del ciclo de vida de los estudiantes (o ex-estudiantes) de la institución.

Al igual que en el caso de los atributos para establecer las alternativas primero se deben recolectar los datos pertinentes (esto se lleva a cabo en el paso siguiente de la metodología).

4.4.1.3 APROXIMACIÓN ANALÍTICA Y REQUERIMIENTOS DE DATOS. (PASO 1.3)

Aproximación analítica En este paso se definen los modelos que se desarrollarán, adaptarán o adoptarán e implementarán para completar la construcción de la estructura del problema de decisión planteado. Como el problema es complejo y presenta muchos factores que pueden incidir en su solución, sería adecuado y por ende recomendable emplear algún modelo o técnica de reducción de dimensionalidad si es necesario, según la situación concreta en una institución: por ejemplo se tienen decenas o cientos de factores o variables a considerar.

Por otra parte es altamente recomendable aplicar técnicas o métodos de aprendizaje supervisado o no supervisado para asignar etiquetas de *DESERTOR* o *NO DESERTOR* a un estudiante en un periodo dado, considerando su registro de ciclo de vida. Pueden aplicarse métodos como: Clasificación Bayesiana, Árboles de decisión, Máquinas de vectores de soporte, entre otros.

Toda vez que se apliquen las etiquetas a todos los estudiantes en todos los pe-

riodos registrados de su registro de ciclo de vida el paso siguiente consiste en aplicar técnicas de reducción de dimensionalidad para identificar cuales son las características mas relevantes que permiten distinguir entre las dos clases mencionadas en el párrafo anterior. Aquí puede ser útil aplicar técnicas estadísticas como correlación, análisis de componentes principales, o técnicas mas sofisticadas como conjuntos de frontera imprecisa (Rough Sets), por citar algunos ejemplos.

Finalmente el problema de decisión multicriterio definido anteriormente tiene en común con estos modelos señalados los datos del registro del ciclo de vida de los estudiantes, pero además requiere de datos específicos que sólo puede suministrar el DM y estos son los que conforman su modelo de preferencias, los mismos se obtienen en un proceso denominando elicitación de preferencias, en el interactúan un analista de la decisión y el DM para plasmar la subjetividad de este último en el método MCDM que se emplee, para elaborar la recomendación final para el DM, en este caso los estudiantes que son identificados como desertores y los que no son identificados como desertores en cada periodo, pero a diferencia de las etiquetas aplicadas por los métodos de ML aquí juega un papel fundamental las preferencias del DM.

Requerimientos de datos Los requerimientos de datos están determinados en gran parte por los métodos analíticos y de apoyo a la decisión seleccionados, y principalmente se obtienen del registro del ciclo de vida del estudiante. Estos datos deben existir física o electrónicamente y deben ser extraídos, cargados en la herramienta de solución, validados, transformados para su posterior utilización en los modelos seleccionados.

Los requerimientos también contienen los datos que conforman el modelo de preferencias del DM asociado al método MCDM seleccionado. Estos datos, por lo general, no existen o no están disponibles en físico o electrónico. Solo se encuentran en la mente del DM y son de caracter subjetivo, por ello para su extracción se requiere de un elaborado, complicado y costoso, al menos en tiempo, proceso de elicitación. Una vez que se obtienen deben ser validados para garantizar su consistencia con

las preferencias del DM, que sólo existen como un sistema de expectativas, juicios, relaciones, etc. en el cerebro del DM.

Toda vez que se han definido los requerimientos de los datos e identificadas las fuentes que proporcionarán los mismos se pasa a la extracción, carga y transformación de estos, lo cual se realiza bajo las recomendaciones del paso siguiente.

4.4.2 CONSTRUCCIÓN DE LA INSTANCIA DE DATOS DEL PROBLEMA. PASO 2

El propósito de este paso es la construcción de las alternativas y la definición y formulación de los atributos y sus escalas. Para dar cumplimiento al mismo es preciso primero llevar a cabo la adquisición, limpieza, transformación y exploración de datos.

Este paso abarca las siguientes fases de las metodologías mencionadas:

MGCD-IBM (Metodología General de la Ciencia de Datos propuesta por IBM): Se aplica para la definición y selección de atributos.(Adquisición, limpieza, validación, y transformación de los datos, Análisis exploratorio de los datos, Pre-procesamiento de los datos, Modelación, Evaluación de modelos.

En este paso no se obtienen los datos relacionados con la información preferencial del DM, esto se llevará a cabo hasta el paso siguiente (Paso 3).

MGADM (Metodología General de Apoyo a la Decisión Multicriterio): Se aplica para la construcción de las escalas de los atributos y las alternativas (Diseño)

4.4.2.1 DEFINICIÓN Y SELECCIÓN DE ATRIBUTOS. PASO 2.1

En base a los resultados de la fase de diagnóstico y el *Paso 1* se procede a extraer los datos de las fuentes identificadas. Básicamente los datos que se extraen están vinculados a la hoja de vida del estudiante. Una vez extraídos se cargan en la plataforma seleccionada para el procesamiento y transformación de los mismos. Sigue un análisis exploratorio de los datos, en el que principalmente se calculan estadísticas descriptivas, así como dependencias, correlaciones y visualizaciones para profundizar en el conocimiento de los datos y definir los atributos que puedan ser extraíbles a partir de estos. Debe verificarse de alguna manera que el conjunto de atributos sea exhaustivo, minimal e independiente Bell *et al.* (1977). De estas características probablemente la más difícil de verificar es que el conjunto de atributos sea exhaustivo, o sea que representa los objetivos del problema. Precisamente en Bell *et al.* (1977) se describe una metodología para lograr que un conjunto de atributos cumpla con estas características, esta se basa en el principio de descomposición descendiente, que inicia por un objetivo general y lo va descomponiendo en objetivos mas concretos hasta llegar a los atributos en las hojas del árbol desarrollado. Pueden tomarse como base los objetivos mencionados en la sección anterior, los que sean pertinentes, y a partir de estos desarrollar el resto del árbol.

Es altamente deseable que este ejercicio se desarrolle conjuntamente con el DM, o de no ser posible que este se involucre en ciertos momentos donde su intervención es requerida para validar el avance. Esto con el propósito de hacer mas eficiente este proceso, ya que mientras más se involucre el DM menor será el chance de que al finalizar el proceso el DM rechace el árbol desarrollado y se tenga que rehacer todo de nuevo.

4.4.2.2 CONSTRUCCIÓN DE LAS ESCALAS DE PREFERENCIA DE LOS ATRIBUTOS. PASO 2.2

Una vez que se han definido los atributos, se procede a definir las escalas de preferencias o dominios para cada uno de estos. Esta es una actividad en la que también debe participar el DM, pero no se requiere que se involucre tanto como en el paso anterior, de hecho, el nivel de involucración y el tiempo que deba invertir el DM en este proceso depende en gran medida del conocimiento del negocio por parte del analista y de su experiencia profesional. Este es un paso crucial en el desarrollo del modelo, por ello es importante aplicar análisis de sensibilidad para las discretizaciones realizadas, a fin de detectar cuales niveles de cuales escalas son mas sensibles para aplicarles a estos procedimientos de verificación mas sofisticados, asegurando que los niveles obtenidos concuerden con las preferencias del tomador de decisiones y cuyo significado sea comprendido por el DM.

Es recomendable implementar varios experimentos pequeños para validar tanto los atributos como las escalas asociadas, simplemente construyendo casos frontera de estudiantes desertores y no desertores y así poder observar como los interpreta el DM y las observaciones o recomendaciones que de estos experimentos se deriven para mejorar la asertividad del modelo.

4.4.2.3 ASIGNACIÓN DE ETIQUETAS DE *DESERTOR* o *NO DESERTOR* A CADA ALTERNATIVA. PASO 2.3

Una vez que se tienen los datos recolectados del registro del ciclo de vida de los estudiantes se pasa a establecer en que periodos del mismo han desertado, para luego, aplicando la definición de desertor asignarles la etiqueta correspondiente en el periodo. Si la definición es mas compleja se pueden aplicar pasos subsiguientes para asignar la etiqueta correspondiente (por ejemplo, en Colombia se considera el concepto de *Desertor por período*, como se había mencionado, que requiere no solo

verificar si un estudiante deserta en un periodo sino en dos periodos consecutivos.

Una vez asignadas las etiquetas (que deben definirse por el DM, por ejemplo podrían ser *No desertor*, *Desertor por periodo*, *Desertor definitivo*) se pasa a validarlas, y se discute el resultado con el DM quien puede sugerir ajustes o cambios en la asignación, mismos que deben implementarse en este paso.

4.4.2.4 CARACTERIZACIÓN DE DESERTORES Y REDUCCIÓN DE ATRIBUTOS.PASO 2.4

Se aplica un conjunto de métodos (mayormente definidos en el paso de la aproximación analítica) para obtener una caracterización de los desertores y no desertores, identificando también cuales atributos hacen aportes mas significativos a esta caracterización.

Independientemente de que el conjunto de atributos cumple con las tres características descritas arriba, los datos disponibles pueden cambiar el impacto de estos en la composición del atributo que establece si un estudiante es desertor o no, ya sea porque un conjunto de atributos no aporte de manera significativa en la distribución de esta característica, o porque existan atributos dependientes o correlacionados, u otras evidencias que identifican a un conjunto de atributos como supérfluos. Es por ello, que se aplican una o varias técnicas para reducir la cantidad de atributos que se emplearán en la caracterización de los desertores. Para esto también pueden utilizarse los resultados de la aplicación de los algoritmos de ML.

A partir de los resultados obtenidos se establece un nuevo conjunto de atributos, que puede contener nuevos atributos que resulten de combinar algunos de los existentes o de calculaciones realizadas a partir de atributos existentes, se eliminan los redundantes y los que tienen una aportación no significativa en la caracterización de los desertores (no desertores).

Ideas para maximizar la robustez de las clasificaciones Como resultado de la aplicación de estos algoritmos se obtiene: Sea F_i un algoritmo de aprendizaje supervisado, P_{ik} un conjunto asignación de valores de los parámetros de ese algoritmo, y C_{jp} los valores de los atributos seleccionados que representan a un estudiante j en un período p de su ciclo de vida en particular. Sea D_{jp} la etiqueta asociada al estudiante en cuestión j en el periodo p , donde $D_{jp} \in Desertor, NoDesertor, \dots otras categorías$, entonces

$$F_i(P_{ik}, C_{jp}, D_{jp}) = S_{ikjp} \quad (4.1)$$

caracterización del estudiante j en el período p , por el algoritmo i , para una asignación de parámetros k .

$$S_{kjp} = \bigcap S_{ikjp} \quad (4.2)$$

conjunto de características fundamentales del estudiante j en el periodo p , para la asignación de parámetros k (puede ser vacío).

Los conjuntos S_{kjp} pueden emplearse para definir un proceso de optimización sobre un conjunto de algoritmos de ML que maximice la robustez de los resultados de las clasificaciones a partir del ajuste de parámetros e hiperparámetros de esos algoritmos, pero también puede incluirse la definición de las categorías de las escalas de los atributos, o incluso los atributos incluidos en el modelo. El objetivo puede formularse como:

$$maximizar \sum |S_{kjp}| \quad (4.3)$$

Las variables de decisión serían los parámetros de los algoritmos aplicados. Las restricciones podrían estar relacionadas con el objetivo, por ejemplo estableciendo una cota mínima para el objetivo.

Podría agregarse un segundo objetivo que busque minimizar la cantidad de atributos empleados en las clasificaciones.

La implementación formal de estas ideas se recomienda como trabajo futuro, por su complejidad y el tiempo requerido de cómputo aún para pequeñas instancias.

4.4.3 FORMULACIÓN DE LA RECOMENDACIÓN. (PASO 3)

Toda vez que ya han quedado establecidos los elementos del problema de decisión, los datos cargados y listos para usarse, es el momento de aplicar un método de apoyo a la decisión multicriterio para resolver el problema de decisión y la elaboración de una recomendación para la decisión. Justamente ese es el propósito de este paso.

Las fases de las metodologías base que se implementan en ese paso son:

MGADM:(Metodología General de Apoyo a la Decisión Multicriterio): Se aplican para resolver el problema de decisión abordado y apoyar en la generación de la recomendación para apoyar la toma de decisiones del DM.(Selección, Implementación.).

MGCD-IBM (Metodología General de la Ciencia de Datos propuesta por IBM): Retroalimentación de la recomendación y realización de ajustes de ser necesario. (Implementación, retroalimentación).

4.4.3.1 SOLUCIÓN AL PROBLEMA DE DECISIÓN. (PASO 3.1)

En este paso se soluciona el problema de decisión para lo cual se deben realizar las siguientes acciones:

1. **Elegir el método de apoyo a la decisión multicriterio** que será empleado: este dependerá del tipo de problema de decisión (*jerarquización, selección, clasificación*), y de la clase de problema de decisión *Bajo certeza, bajo incertidumbre, bajo estricta incertidumbre*, entre otros elementos. También puede ser requerido, por la complejidad del problema que se requiera adaptar un método existente o crear un nuevo método.

2. **Construir el modelo de preferencias del DM** acorde al método de decisión elegido o elaborado.
3. **Aplicar el método de apoyo a la decisión multicriterio** a la instancia de datos construida para llevar a cabo la caracterización de los desertores.
4. **Validar los resultados de la aplicación del método con el DM, y de ser necesario ajustarlo y volver a aplicarlo hasta que el DM este satisfecho.**

El problema que se plantea resolver es un problema de decisión bajo certeza y es del tipo de ordenamiento o evaluación multicriterio, por lo que los métodos que se apliquen, adapten o desarrollen deben enfocarse en este tipo de problemas. Entre los mas populares se encuentran: *la familia de ELECTRE TRI*, *los métodos AHP*, *recientemente métodos basados en PROMETHEE* entre otros, también es posible emplear combinaciones de métodos para solucionar problemas complejos.

En esta metodología se propone emplear una combinación de tres métodos para resolver el problema de evaluación multicriterio abordado:

- **AHP**: Para la generación de pesos o importancia relativa de los atributos.
- **ROUGH SETS**: Para la caracterización multicriterio de las categorías de deserción y apoyar la definición de perfiles que definan las categorias de permanencia o deserción escogidas.
- **ELECTRE TRI**: Para realizar la asignación de alternativas a categorías.

Asignación de pesos o importancia relativa de los atributos. Uno de los retos mas importantes en la correcta aplicación de los métodos de apoyo a la decisión bajo certeza es el establecimiento de los pesos de los atributos o importancia relativa de los mismos, información que debe ser suministrada por el DM. La dificultad en que estos valores representen realmente las preferencias del DM estan perfectamente documentadas y discutidas en Roy; Zopounidis y Doumpos (2002); French (1986).

Saaty expone claramente porqué resulta tan difícil para los seres humanos y en particular los DM establecer la importancia relativa de los atributos Saaty (2004): se requiere establecer una referencia y en base a esta cuantificar la preferencia en términos de diferencias, lo que es extremadamente difícil dado que los números de forma natural y habitual los interpretamos de forma ordinal, mientras la importancia relativa requiere hacer una interpretación cardinal. A continuación, se ilustra la diferencia con ejemplo:

Por ejemplo "3 es mayor que 4" es algo que todos entendemos, sin requerir más explicación o cálculos, por otra parte si se menciona que en la India se reportan cerca de 42000 casos diarios de nuevos contagios por COVID-19, mientras que en Sudáfrica se tienen 16000 (datos del 17 de julio del 2021, tomados de internet), intuitivamente sabemos que la India tiene mayor población que Sudafrica, pero ¿es mayor la tasa de infección en la India o Sudafrica? y ¿cuán grande es la diferencia?. Para responder estas preguntas podríamos establecer una medida que estandarice el número de casos, digamos por cada 1000 habitantes para poder hacer una comparación objetiva:

India tasa de nuevos infectados por cada 1000 habitantes:

$$\frac{42000}{1366417754} = 0.00003 \quad (4.4)$$

Sudáfrica tasa de nuevos infectados por cada 1000 habitantes: $\frac{16000}{59622000} = 0.0003$

Como se puede observar la tasa de infección de Sudáfrica es mayor que en la India y además es 10x veces mayor!, esto a pesar de que tenemos el conocimiento de que la población de la India es mayor que la de Sudáfrica. Y son elementos objetivos, cuando se trata de preferencias se complica aún más la comparación pues se trata de juicios de valor, que no tienen necesariamente una interpretación objetiva o cuantitativa.

AHP ayuda al DM a reflexionar sobre sus preferencias y expresarlas cuantitativamente por medio de una escala de proporciones, brindando además mecanismos para verificar la consistencia de la información dada por el DM, por ello es un método muy popular, aunque se recomienda que se utilice en conjunto con un analista de

la decisión.

Vale la pena mencionar que existen ciertas diferencias entre el concepto de pesos empleado en AHP, por su naturaleza compensatoria y los pesos o importancia relativa empleados en los métodos ELECTRE Figueira *et al.* (2016), sin embargo aún así resulta mas efectivo emplear AHP que solo la intuición para derivar los pesos. Si no se obtienen resultados satisfactorios se recomienda emplear el método SMART Patel *et al.* (2017) para la obtención de los pesos, sin embargo este es un procedimiento mas complejo y requiere de la participación de un analista de la decisión asistiendo al DM durante todo el proceso de elicitación.

Caracterización multicriterio de las categorías de deserción. Los conjuntos aproximados (Rough Sets) Pawlak (2012) ha demostrado ser una valiosa teoría, con sólido formalismo matemático, que es ampliamente utilizada en el análisis de datos, la inteligencia artificial, y el apoyo a la decisión.

En esta propuesta se emplean los Conjuntos aproximados para establecer caracterizaciones minimales de las categoría de deserción o permanencia en base a las descripciones obtenidas mediante la aplicación de algoritmos de ML en el paso anterior (Paso 2).

Los rough sets se emplean para generar las reglas de decisión que caracterizarán cada categoría de decisión, reglas que pueden ser empleadas para definir los perfiles que identifican las categorías de deserción y permanencia estudiantil. Un conocido algoritmo para la generación de reglas es el LEM2 Grzymala-Busse (1992), que es el sugerido se emplee en esta metodología.

LEM2 tiene la característica de que se puede ejecutar estableciendo un error máximo buscando minimizar la cantidad de reglas generadas, lo que permite investigar la relación entre estas dos medidas para obtener un compromiso aceptable entre ambas.

Una vez que se aplica LEM2 y se obtiene un número manejable de reglas con un

error de clasificación aceptable, se procede a definir aquellas reglas que servirán para separar las categorías del problema de decisión, o perfiles como se les denomina en el método ELECTRE TRI. Es importante que estos perfiles se escojan en conjunto con el DM.

Asignación de alternativas a categorías de deserción o permanencia A diferencia de lo que se planteó en la asignación de etiquetas de deserción o permanencia empleando métodos de ML, en ese paso se incorporan las preferencias del DM para definir las categorías de deserción o permanencia, alineando así estas con los objetivos estratégicos de las instituciones.

ELECTRE TRI es un popular método para resolver complejos problemas de evaluación multicriterio Mousseau y Slowinski (1998), un método que sigue vigente y al cual se le agregan nuevas variantes como ELECTRE TRI B, ELECTRE TRI C Bouyssou y Marchant (2015).

Una vez que se ha aplicado y validados los resultados con el DM, llega el momento de poner el método en producción. Hay que tener en cuenta que si cambia la visión estratégica de la institución o cambia el DM se debe aplicar nuevamente este paso, y el anterior.

4.4.3.2 IMPLEMENTACIÓN Y RETROALIMENTACIÓN. (PASO 3.2)

Este es el paso final de la metodología, en este paso se deben tomar decisiones tácticas y operativas para implementar las herramientas de apoyo a la decisión en el sistema de gestión y permanencia. Probablemente se tengan que definir y solucionar otros problemas de decisión asociados a la elaboración de un plan de acción, planeación de recursos, capacitación del personal, etc.

Es un paso muy importante pues si no se realiza adecuadamente, no importa

todo lo hecho anteriormente, simplemente puede fracasar el proceso.

Es por ello también importante diseñar mecanismos de retroalimentación eficientes en cuanto a la decisión tomada e implementada para actualizar las acciones planificadas, así como el plan de acción de ser necesario para que los efectos positivos de la decisión perduren en el tiempo.

CAPÍTULO 5

APLICACIÓN DE LA METODOLOGÍA A UN CASO DE ESTUDIO

5.1 DESCRIPCIÓN GENERAL DE LA UNIVERSIDAD SIMÓN BOLÍVAR USB

La Universidad Simón Bolívar fue fundada el 15 de noviembre de 1972 (Resolución 1318 del 15 de noviembre de 1972 de la Gobernación del Atlántico, por medio de la cual se concede la personería Jurídica), como Institución Universitaria, con el nombre de Corporación Educativa Mayor del Desarrollo Simón Bolívar, con cinco (5) programas profesionales: Derecho, Trabajo Social, Sociología, Economía, Licenciatura en Ciencias Sociales con ochenta y tres (83) estudiantes, pertenecientes un 98 % a los estratos sociales mas bajos (1, 2 y 3). Cómo lo muestra la figura 5.1: Recibe su reconocimiento como universidad en el año 2015 con catorce (14) programas profesionales de pregrado, veintinueve (29) especializaciones de posgrado y un total de 9285 estudiantes. Recibe acreditación de alta calidad en el año 2016, considerada una de las mas altas distinciones a nivel de Universidades, con un total



FIGURA 5.1: Paso del tiempo Universidad Simón Bolívar. Colombia

de 9805 estudiantes. En la actualidad, la universidad cuenta con 5 facultades:

- Administración y negocios
- Ingenierías
- Ciencias Jurídicas y Sociales
- Ciencias de la salud
- Ciencias básicas y biomédicas

En dichas facultades se encuentran distribuidos 21 programas profesionales y 12519 estudiantes de acuerdo a la figura 5.2 y 59 programas de posgrado entre especializaciones, maestrías y doctorados, con un total de 1134 estudiantes.

Actualmente la universidad cuenta con 980 profesores de los cuales 383 son profesores de planta, distribuidos en los diferentes programas que ofrece. Se cuentan con 9 programas cuentan con Acreditación de Alta Calidad, de los cuales 5 están reacreditados, cumpliendo con su compromiso de trabajar en la calidad académica de manera constante y permanente, así como 4 de los programas de pregrado cuentan con doble titulación en universidades extranjeras de Francia, Perú, Chile y México. Pertenece a 116 redes académicas y de investigación a nivel nacional e internacional, tales como la Red iberoamericana en Ciencias Sociales con enfoque de género, Red

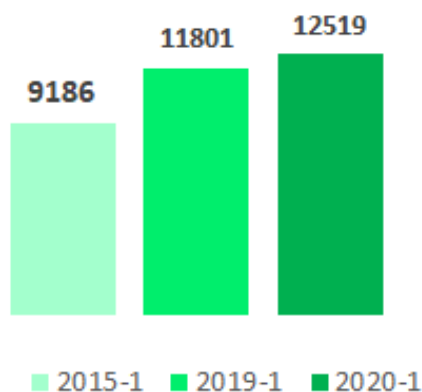


FIGURA 5.2: Matriculas Pregrado

Iberoamericana de investigadores en Salud Renal y enfermedades Crónicas, Red iberoamericana de Justicia Alternativa, tradicional y restaurativa, Res iberoamericana de Investigadores en Análisis de los sistemas regionales de Ciencia, tecnología e innovación en Latinoamérica (ASRIL), Red Iberoamericana de investigación de estudios interculturales e Interdisciplinarios, Red de avances académicos e investigativos en instrumentación quirúrgica y salud, entre otras. La Universidad Simón Bolívar ha obtenido reconocimientos en una serie de ranking a nivel nacional e internacional ubicándose entre una de las tres instituciones de educación superior con mayor impacto social en Colombia. Así lo confirman los resultados generales del ranking mundial de Impacto Social realizado por Times Higher Education (THE), en su edición 2020. El ranking evalúa el desempeño global de las universidades a nivel mundial, con respecto a los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas (ONU). Este estudio utiliza indicadores que son calibrados para proporcionar comparaciones integrales y equilibradas en tres áreas: investigación, divulgación y administración. Igualmente el Ranking Times Higher Education destaca a Unisimón entre las mejores universidades de Latinoamérica. La Universidad Simón Bolívar es una de las mejores de Latinoamérica, de acuerdo con la última edición del listado de la consultora británica en educación superior Quacquarelli Symonds (QS). Unisimón se ubicó entre el 74 % superior en el QS World University Rankings de Latinoamérica, ocupando la posición 301 de 410 instituciones de educación superior de Brasil, México, Colombia,

Argentina, Perú, Ecuador, Chile, Venezuela, Cuba y Costa Rica que lograron entrar a este selecto grupo. de la misma manera, se obtuvo reconocimiento, por segundo año consecutivo, con el Premio Colombiano a las Buenas Prácticas Laborales, en la categoría entidades sin ánimo de lucro –ESALES y Universidades. Unisimón fue evaluada bajo el Modelo de Excelencia de Sostenibilidad en Buenas Prácticas Laborales (MES+BPL), en análisis integral de gestión organizacional, basado en los Estándares Internacionales de Sostenibilidad; y se destacó por promover un modelo de alta exigencia en la gestión de la responsabilidad como parte de las estrategias de sostenibilidad, alineada a las mejores prácticas nacionales e internacionales.

En la gráfica 5.3 se muestra los reconocimientos al pasar de los años:

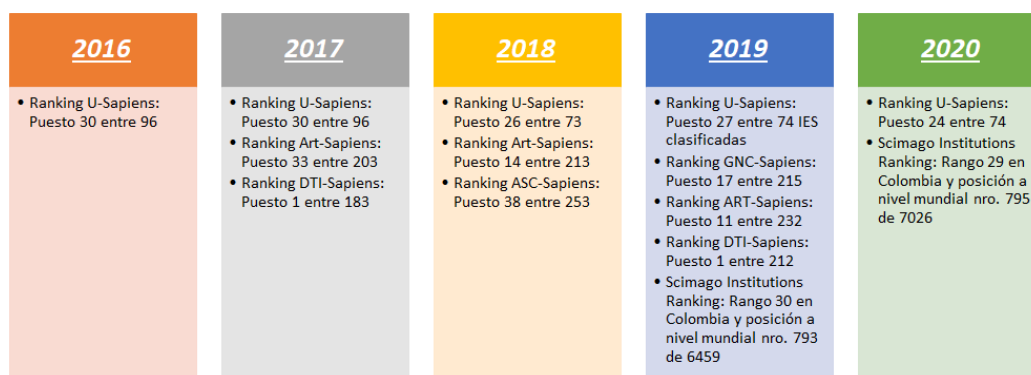


FIGURA 5.3: Ranking Universidad Simón Bolívar

La universidad cuenta con 85 profesores con Doctorado, 572 con maestría, 176 con especializaciones médico quirúrgica, 102 con especialización y 45 con estudios profesionales, para apoyar a su población estudiantil y los diferentes proyectos que se gestan en cada uno de los ejes misionales de Docencia, Investigación y Extensión y proyección Social, cumpliendo con los objetivos planteados en el plan estratégico de desarrollo institucional y de igual manera con los planes de acción anuales.

Por último, de acuerdo al **Informe de rendición de cuentas 2020** disponible en de Planeación Universidad Simón Bolívar (2021) tenemos la siguiente infografía:

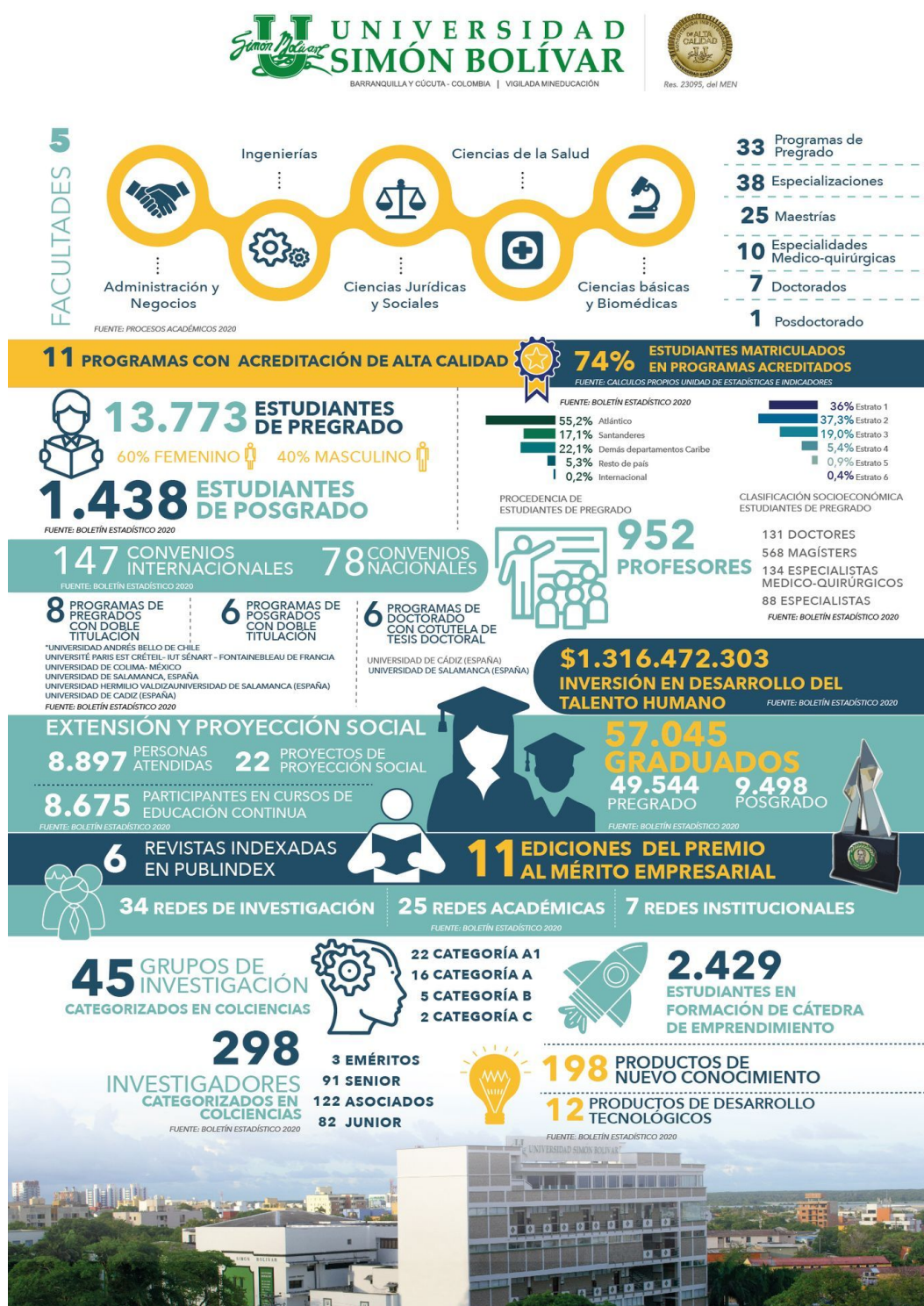


FIGURA 5.4: Universidad Simón Bolívar en Cifras. 2020

5.1.0.1 DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD SIMÓN BOLÍVAR. BARRANQUILLA, COLOMBIA

La Universidad Simón Bolívar sede Barranquilla, como todas las Universidades en Colombia, mide las cifras de deserción a través del SPADIES (2020) Sistema para la Prevención de la Deserción en la Educación Superior (SPADIES), sin embargo, maneja sus estadísticas internas, dónde se clasifican los porcentajes de deserción de los estudiantes de acuerdo a las categorías de Académicas, Individual, Económica y la Institucional, por cada Facultad, por período.

Desde el análisis de la deserción por período entre los años 2015 y 2018, se evidencian el número de estudiantes en cada categoría, reflejada en la gráfica 5.5 a continuación:

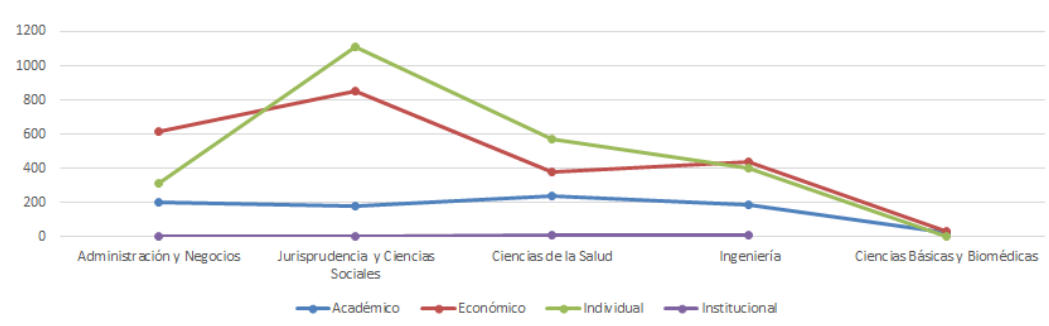


FIGURA 5.5: Porcentaje de estudiantes en las Categorías de Deserción

La figura resalta que las variables mas recurrentes en la deserción, son las relacionadas con la categoría individual en la mayoría de los programas asociados a cada facultad, seguida por la Económica, excepto en el caso de la Facultad de Administración y Negocios, dónde el mayor porcentaje, se encuentra en la categoría Económica

De esta manera, se realizan al interior de la universidad, acciones independientes para apoyar al estudiante en cada uno de las categorías y poder continuar con sus estudios hasta la graduación, sin embargo, solo se atienden a aquellos, quienes son detectados o quienes voluntariamente solicitan los servicios a sus programas

académicos o a las unidades administrativas encargadas. De esta manera, a pesar de tener la información semestre a semestre de las razones de abandono, la Universidad aún no tiene una manera de detección temprana, para evitar la deserción de los estudiantes de manera efectiva.

5.2 FASE DE DIAGNÓSTICO

En esta fase se trabajará básicamente lo expuesto en la imagen 5.6 aplicada para la Universidad Simón Bolívar:

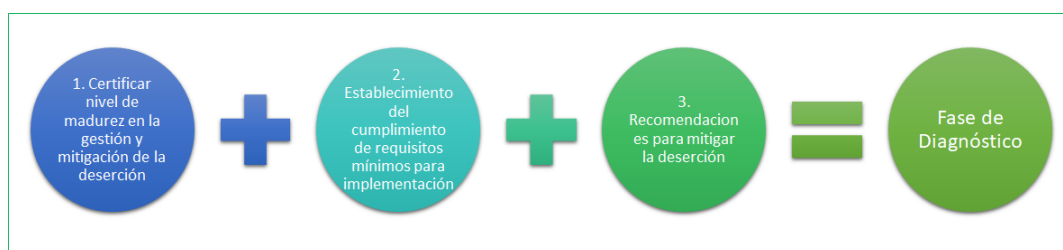


FIGURA 5.6: Fase Diagnóstico

5.2.1 APLICACIÓN DE LA FASE DE DIAGNÓSTICO A LA USB

En la Universidad Simón Bolívar, desde el análisis realizado en cuanto al diagnóstico de la capacidad institucional, desde la revisión de documentación, página web, observación directa de procesos y actividades en las diferentes áreas de la Universidad que apoyan la permanencia estudiantil y trabajan en pro a la mitigación de la deserción, se puede decir que el problema está en un nivel de gestión de interés en **En reconocimiento** de acuerdo a lo planteado en la metodología. Frente al modelo de gestión de permanencia y graduación estudiantil para IES, propuesto por el MEN, se encuentran fortalezas y aspectos por mejorar, los cuales se describen a continuación, desde cada uno de los componentes:

- Sensibilización - Socialización
- Cultura de información
- Mejoramiento de la calidad
- Trabajo en conjunto con Instituciones de educación media (IEM)
- Programa de apoyo a estudiantes
- Compromiso con el núcleo familiar
- Gestión de recursos
- Trabajo colaborativo

De acuerdo a lo establecido en la Guía para la implementación del modelo de gestión de permanencia y graduación estudiantil en Instituciones de educación Superior (IES) Ministerio de Educación Nacional Colombiano (2015) así:

1. Sensibilización - Socialización 4.2: En cuanto a este componente la universidad debe trabajar en:

- Normatividad: Se deben crear o actualizar documentos de:
 - Política de permanencia
 - Procedimientos en el sistema de gestión de calidad (SGC)
 - Reglamento estudiantil. Parágrafos de obligatoriedad del uso de las estrategias de acompañamiento.
- Plan de comunicaciones: Desarrollar:
 - Campañas de motivación para el uso de estrategias Institucionales de acompañamiento para docentes y estudiantes. (Cultura De calidad Académica)
 - Sitio web

- compromiso de Participación activa de los docentes en las estrategias de permanencia de las IES
 - Informes: Generar informes relacionados con:
 - Gestión frente a la permanencia con medición de impacto de estrategias institucionales aplicadas.
 - Gestión frente a la deserción con correlación de variables incidentes.
2. Cultura de la Información 4.3: Desarrollar, actualizar y afinar desde el sistema de información institucional:
- Caracterización de estudiantes:
 - Digitalización de pruebas.
 - Caracterización socioeconómica y académica de Estudiantes
 - Correlación de variables (Académicas, Financieras y Psicológicas).
 - Definición de indicadores de impacto de uso de estrategias institucionales
 - Optimización de Sistemas de información nacionales de prevención a la deserción en instituciones de educación superior SPADIES y Sistema Nacional de Instituciones de educación superior (SNIES)
 - Uso de SPADIES por todos los directores de la Universidad. Capacitaciones incluidas en plan de acción del Dpto de Planeación 2017.
 - Observatorio de permanencia académica
 - SNIES
 - Seguimiento estudiantil académico:
 - Alertas tempranas
 - Seguimiento y Medición de impacto : Módulos: Remisiones por docente (rutas), asistencia, tutorías, monitorias, asesorías, consejerías)
 - Reportes: Cursos críticos, repitencia, reprobados, Icetex, Pilo Paga, VCA

- Hoja de vida estudiantil
 - Costos de deserción
3. Mejoramiento de la calidad académica: Se describen a continuación acciones en desarrollo (**) y acciones por aplicar para fortalecer el componente:

- Fortalecimiento de la gestión curricular
 - Macro proyecto de Gestión Curricular **
- Plan de formación Docente
 - Capacitaciones de tutores exitosos
 - Modelo Basado en Evidencia
 - Estilos de aprendizaje
 - Informe de seguimiento y de impacto
 - OVAS para docentes
- Evaluación Docente **
 - Evaluación que contenga la medición de competencias funcionales y de comportamiento de los docentes, que estimule el crecimiento personal y el desempeño profesional
- Metodología de enseñanza basada en estilos de aprendizaje
 - Guías, OVAS, o materiales educativos que orienten a los docentes y estudiantes sobre estrategias, dinámicas y herramientas, según los estilos de aprendizaje.
 - Medición de elaboración y aplicación de Guías.
- Aulas virtuales de aprendizaje del 100 % de los cursos ofertados
 - Moodle **
 - Biblioteca Digital **

Continuando con el Mejoramiento, tenemos además 4.5:

- Objetos virtuales de aprendizaje (OVA)

- Identificación de cursos críticos y construcción de objetos virtuales de aprendizaje OVA de los mismos.
- Construcción de OVAS para motivación, manejo de ansiedad ante exámenes, preparación para parciales, concentración y memoria, pilates cerebral, etc.
- Monitorias
 - Se tiene plan de capacitación de monitores **
 - Se cuenta con procedimiento de monitores que incluye: horas asignadas, cursos asignados, criterios de selección, descuentos asignados. PIEA propone mejorar la oferta de descuentos ofrecidos a los monitores **
 - Informe de seguimiento y de impacto de cursos vs rendimiento
 - Grupos de estudio: Con estudiantes talentos y destacados en cursos críticos, se crean los grupos de estudio entre estudiantes.
- Tutorías:
 - Preparación de docentes para tutoría exitosa en cursos identificados críticos
 - Impacto de la estrategia. Rendimiento académico en curso vs tutoría
 - Identificación de cursos críticos
 - Asignación de horas docentes para los cursos críticos detectados en los Programas.
- Consejerías Académicas:
 - Servicios de: orientación a los estudiantes y padres de familia.
 - Articulación con los diferentes servicios y áreas que contribuyen a la permanencia estudiantil **
 - Identificación del nivel de satisfacción de los estudiantes con el programa académico al cual están adscritos y con los servicios que les ofrece la institución.

- Programas de apoyo grupal
- Talleres y tips mensuales de temas identificados como críticos.

4. Trabajo en grupo con las Instituciones de Educación Media (IEM) 4.6

- Nivelación y refuerzo académico
 - Definición de cursos nivelatorios pre semestre y de refuerzos durante el semestre.**
 - Cursos de nivelación inicial en Matemáticas y Lectoescritura y Nivelación en segunda Lengua
- Articulación y Cualificación docente IEM
 - Articulación con docentes de IEM, construcción de convenios para capacitación docentes de IEM desde los posgrados ofrecidos en la USB relacionados con la docencia o crear programas de formación docentes IEM específica con Colegios quienes mas nro de estudiantes se tienen matriculados. (Estudio de admisiones). Diplomados, seminarios, etc.
- Acompañamiento para la vinculación a la educación superior
 - La dirección de Admisión participa activamente en el proceso de visita a las IEM para captar estudiantes para ingresar en los programas ofertados por la Universidad. Orientación Vocacional.**
 - Realizar Feria de Orientación Vocacional.
 - Portafolio de servicios para los estudiantes de IEM: través de Pruebas de orientación vocacional Online, charlas básicas de orientación vocacional, entre otros.**

5. Programa de apoyo para los estudiantes 4.7

- FORTALECIMIENTO DEL BIENESTAR INSTITUCIONAL
 - Programas que promuevan Actividades de Salud, Desarrollo Humano, Cultura y Deporte, garantizando la formación integral del estudiante,

la inclusión educativa y la capacidad del estudiante para asociar los programas de formación con su proyecto de vida.

- Orientación educativa
- Acompañamiento para el desarrollo de competencias para la vida
- Promoción socioeconómica
- Fomento de la actividad física, el deporte y la recreación
- Expresión cultural y artística
- Promoción de la salud integral y autocuidado
- Cultura institucional y ciudadana
- Caracterización de la población: Docentes, estudiantes, egresados, padres
- Proceso de Inducción: Integración de los padres a los procesos de inducción
- Orientación frente a la crisis de carrera y adaptación:
 - Programas de Orientación Socio ocupacional
 - Cursos de educación financiera
 - Proyecto de vida
 - Cursos profesionalizantes nivel básico
- Orientación para el trabajo de grado y las prácticas laborales
 - Ruta de acompañamiento a estudiantes para trabajos de grado y la incursión al mundo laboral.
 - Informe de evaluación de impacto de estrategias propuestas
- Servicios para el ingreso al mundo laboral: Un programa o plan de servicios para el ingreso al mundo laboral**
- Observatorio de graduados: Observatorio de egresados, como insumo para ofrecer servicios educativos o replanteamiento de los cursos ofrecidos al interior de los programas de la Universidad

6. Compromiso con el núcleo familiar

- Canales de comunicación con las familias:

-

Programa acompaña a tu hijo en la U, sin embargo no hay datos confiables de los mismos **

- Institucionalización del encuentro de padres. PIEA
- Sitio web de consulta para padres
- Redes familiares: Red de Padres: sensibiliza a la familia sobre la responsabilidad en la formación del estudiante y brinda herramientas de seguimiento, motivación y apoyo, que ellos pueden adoptar y apropiar.
- Programa de formación familiar: Talleres de Formación para padres
- Apoyo al emprendimiento familiar:
 - Promover bienestar en el hogar.
 - Estudiar y solucionar las problemáticas socioeconómicas.
 - Aumentar la calidad de vida y tiempo destinado a compartir en familia.
 - Vincular a la familia en el proceso adaptativo y de innovación del estudiante.
 - Desarrollar una mentalidad emprendedora

7. Gestión de recursos:

- Servicios de Financiamiento: Aumentar servicios de financiamientos ofrecidos **
- Fondo de becas: Fortalecimiento programa de becas institucionales
- Convenios de colaboración para el apoyo a estudiantes: Plan padrino

8. Trabajo colaborativo entre IES:

- Alianzas y transferencia de conocimientos: Consultorías, Investigación u otras actividades en conjunto con otras IES para intercambio de experiencias exitosas

- Redes de conocimiento institucional: Red de fomento a la permanencia y graduación estudiantil
- BENCHMARKING: Análisis de experiencias significativas de otras instituciones, tanto nacionales como internacionales, le permite contar con el conocimiento del avance y del impacto que tiene el programa de permanencia de la institución frente a otros, en términos de factores, variables, estrategias y acciones comunes

5.2.2 FORMULACIÓN DE LA RECOMENDACIÓN DE LA FASE DE DIAGNÓSTICO

De acuerdo con la metodología propuesta, el paso siguiente, consiste en establecer las escalas de preferencias para los criterios o atributos de decisión. Se utiliza una escala estándar con cuatro niveles de preferencia:

- *Sin impacto*: Es el nivel mas bajo de preferencia que indica ausencia de una característica correspondiente a un criterio o atributo, o su nivel de implementación aún no genera impacto observable en el sistema.
- *Impacto mínimo*: Este nivel indica que una característica asociada a un criterio o atributo si está presente pero que aun tiene muchas áreas de oportunidad que deben abordarse para que tenga un real impacto positivo en el sistema.
- *Impacto suficiente*: Indica que la característica tiene un nivel de implementación que tiene un impacto positivo en el sistema aunque aun tiene áreas de oportunidad.
- *Impacto importante*: Este nivel indica que la característica tiene un impacto positivo importante en el sistema, y que aunque tiene áreas de oportunidad en su implementación es una caracterísca madura.

Se procede a realizar la selección de los niveles de preferencia de acuerdo a cada criterio y lo encontrado en la Universidad, dando como resultado lo siguiente 5.2.2:

Criterios	Actividad	Indicadores	SIN IMPACTO	IMPACTO MÍNIMO	IMPACTO SUFICIENTE	IMPACTO IMPORTANTE
Gestión y autoevaluación	Proceso de admisión	Identificación de posibles desertores		X		
		Recomendación de acompañamiento para posibles desertores			X	
	Gestión de la permanencia del estudiante	Integrada en el sistema de calidad		X		
		Integrada en el sistema de enseñanza-aprendizaje Evaluación o fortalecimiento de capacidades de estudiantes		X X		
Seguimiento y apoyo académico	Evaluación / Fortalecimiento de capacidades de estudiantes	Se aplican exámenes de diagnóstico		X		
		Fomento y apoyo a grupos de estudio	X			
	Mecanismos de asesoría y tutoría	Capacitación tutores	X			
		Disponibilidad de espacios y recursos para tutorías y asesorías Integración con otras actividades académicas		X X		
Apoyos Económicos	Asesoría, otorgamiento y seguimiento de apoyos	Existen estrategias, políticas para otorgar apoyos		X		
		Existen mecanismos para asignación de apoyos externos a estudiantes		X		
	Evaluación de impacto de apoyos en permanencia y calidad de resultados académicos	Existen mecanismo de seguimiento a los apoyos asignados a los estudiantes. Evaluación del impacto en la permanencia de los apoyos económicos otorgados.	X	X		
		Los resultados de la evaluación del impacto son accesibles y están integrados con el sistema de enseñanza-aprendizaje.	X			
Seguimiento de la situación individual y personal	Seguimiento y actualización de situación	Existen mecanismos simples y efectivos para actualizar la situación individual o personal del estudiante cuando se ocupe Existen mecanismos para dar seguimiento a la situación individual o personal del estudiante cuando se ocupe	X	X		
Cultura de la información	Gestión y autoevaluación	Existen estrategias para la automatización de la gestión de la información relacionada con la gestión de la Permanencia.		X		
		Existen políticas (reglamentos) para la publicación, acceso y protección de la información.		X		
	Seguimiento y apoyo académico	Está implementada la hoja de vida del estudiante.		X		
		Generación de reportes de seguimiento en tiempo real.		X		
		Acceso y disponibilidad a la información.		X		
	Apoyos económicos	Mecanismos simples y flexibles para actualizar la información.		X		
		Esta automatizada la gestión de la información relacionada con apoyos económicos.	X			
	Situación individual o personal	Está integrado al sistema de enseñanza-aprendizaje la información sobre apoyos económicos.	X			
		Existe un sistema, módulo para gestionar, actualizar información relacionada con la situación individual y/o personal del estudiante. La gestión de la información del seguimiento está integrada con otros sistemas informáticos de la institución.	X X			

De acuerdo a lo indicado en la metodología, para poder identificar la categoría del nivel de gestión se tiene se remite a lo definido así:

Perfil 1:

- Dentro de **Gestión y Autoevaluación** se considera: *Gestión de la permanencia*.
 - Integrada en el sistema de enseñanza-aprendizaje: **Impacto Mínimo**
- Dentro de **Seguimiento y apoyo académico** se considera *Mecanismos para asesorías y tutorías*.
 - Disponibilidad de espacios y recursos para tutorías o asesorías: *Impacto mínimo*.
 - Integración con otras actividades académicas: *Impacto mínimo*.

Perfil 2: Este perfil se relaciona directamente con el atributo de **Cultura de la información** en el subatributo *seguimiento y apoyo académico*..

1. Está implementada la hoja de vida del estudiante: *Impacto mínimo*.
2. Generación de reportes de seguimiento en tiempo real: *Impacto mínimo*.
3. Acceso y disponibilidad a la información: *Impacto mínimo*.
4. Mecanismos simples y flexibles para actualizar la información: *Impacto mínimo*.

De acuerdo a la evaluación realizada, la universidad se encuentra en la categoría de *Reconocimiento*, siendo este es el nivel de madurez más bajo en la evaluación del sistema, se identifica porque no existen estrategias, tácticas o mecanismos implementados para mitigar la deserción o de haberlos no tienen un impacto positivo

observable y documentado y en este caso, los mecanismos implementados aún no tienen un impacto documentado

De acuerdo a los resultados obtenidos en el análisis previo, lo encontrado en la institución, no es suficiente para aplicar la fase de implementación, sin embargo, con los datos recopilados en el marco de la investigación se cuentan con ciertos resultados, los cuales son procesados con las diferentes herramientas propuestas en la metodología obteniendo resultados parciales, propios de la población estudiantil de la universidad.

5.3 FASE DE INTERVENCIÓN

5.3.1 DEFINICIÓN Y ESTRUCTURA DEL PROBLEMA DE DECISIÓN(PASO 1)

5.3.1.1 DEFINICIÓN DEL PROBLEMA DE DECISIÓN. (PASO 1.1)

El problema de decisión que se aborda, como se mencionó en la sección anterior consiste en caracterizar a los estudiantes en Desertores o No Desertores acorde a características: Académicas, Económicas, Institucionales e Individuales. Pues estas mismas son las que se toman en cuenta en la USB.

5.3.1.2 ESTRUCTURA DEL PROBLEMA DE DECISIÓN.(PASO 1.2)

La estructura del problema de decisión consiste en determinar los elementos fundamentales del mismo como son:

- Tomador de decisión (que es un rol, pudiendo ser una persona o un grupo de personas), DM por sus siglas en inglés. En este caso el DM sería la Vice rectoría académica de la USB con el comité de apoyo a los procesos de permanencia y deserción (PIEA y Bienestar Universitario), así como los directores de las facultades que representan sus intereses en cada una de ellas. En la Gráfica 5.7 se puede ver el Organigrama interno de la Universidad, en el área de Vice rectoría académica y las coordinaciones, departamentos y unidades administrativas adscritas a ella, que velan por la permanencia y deserción estudiantil, así:



FIGURA 5.7: Organigrama

- Atributos o criterios: Se deben considerar de de acuerdo a lo que permite medir las variables de cuatro Factores y/o dimensiones 5.8:
 1. Académica: Orientación socio ocupacional, Tipo de colegio, Rendimiento académico, Pruebas Nacionales, Profesores, Métodos de estudio o aprendizaje.
 2. Individual: edad, sexo, estado civil, estado de salud, integración social, entorno familiar
 3. Socio-Económica: Estrato social, situación laboral, Dependencia económica, situación laboral de padres, Personas a cargo.

4. Institucional: Normalidad académica, servicios de financiamiento, Recursos universitarios, Apoyo académico, apoyo psicológico.

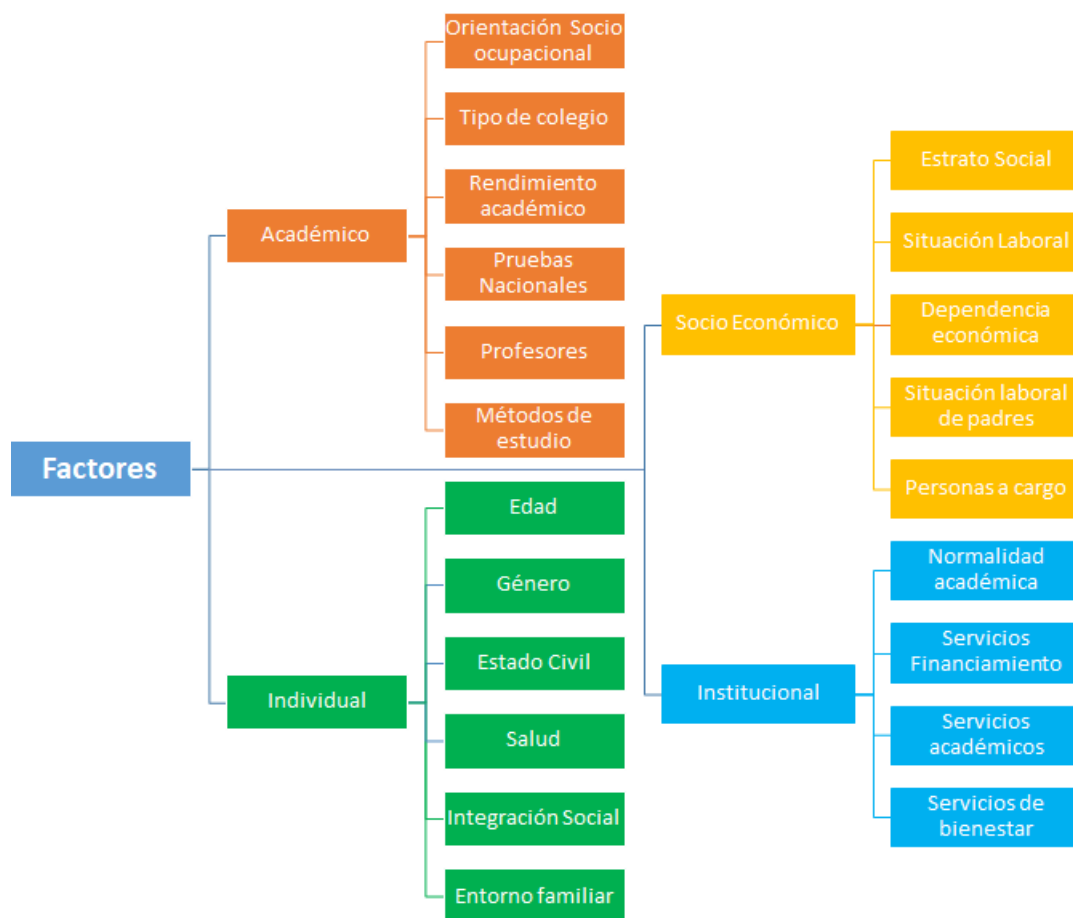


FIGURA 5.8: Factores a considerar para la deserción

- Alternativas: Los estudiantes que han ingresado en la USB y de los cuales se tiene un registro completo de su hoja de vida en el periodo que se considere relevantes para el estudio

5.3.1.3 APROXIMACIÓN ANALÍTICA Y REQUERIMIENTOS DE DATOS. (Paso 1.3)

Se decidió aplicar algoritmos de aprendizaje automatizados para llevar a cabo la caracterización de los factores asociados a las dimensiones de propuestas en la hoja de vida de los estudiantes de manera tal, que los factores que se obtengan sean un conjunto minimal y consideren una mejor y mayor validacion posible para lo posibles desertores

para dar cumplimiento a este paso se decidió utilizar el RapidMiner para aplicar los modelos de Aprendizaje supervisado para caracterizar a los desertores. La decisión de usar RapidMiner se fundamentó principalmente en la facilidad que brinda para preprocesar, grandes y diversos volúmenes de datos, así como la generación automática de modelos de aprendizaje supervisado y su consecuente ajuste para obtener resultados confiables y de calidad. Lo que acorta considerablemente el tiempo para obtener la solución. Para profundizar en la caracterización y generar un conjunto único de reglas reducidas por cada facultad sugiere emplear Rough Sets, o conjuntos de frontera imprecisa por su generalidad y prácticamente no imponen requerimientos fuertes que deben cumplir los datos (como los estadísticos por ejemplo) para garantizar la validez de los resultados. Sin embargo, esta actividad no se llevó a cabo en esta investigación, pues apenas si se realiza una primera etiquetación de los desertores, se necesita a posteriori discutir con las autoridades los casos en los que hay discrepancias con las etiquetas asignadas por la institución para establecer un concepto de desertor en función de los atributos seleccionados y a partir de aquí generar finalmente las reglas de caracterización y clasificación.

En base a estas decisiones los requerimientos de datos para poder aplicar estos métodos son que se tengan registros de los estudiantes en las dimensiones que se consideran y que se puedan construir a partir de estos los atributos que se definieron anteriormente.

5.3.2 CONSTRUCCIÓN DE LA INSTANCIA DE DATOS DEL PROBLEMA (PASO 2)

5.3.2.1 DEFINICIÓN Y SELECCIÓN DE ATRIBUTOS (PASO 2.1)

EXTRACCIÓN DE DATOS: El proceso de obtención de datos, se realiza desde tres insumos, los cuales incluyen los Sistemas de información Institucionales, y reportes solicitados directamente al Departamento de Tecnología de la Universidad y los generados internamente en una base de datos propia para el proyecto. El sistema de información Institucional SIAAF, tal y como se expresa en el **Anexo2** genera una amplia variedad de reportes que pueden ser descargados en diferentes formatos. Para esta investigación se descargaron en formato .csv, mismos que se describen a continuación:

- Matrículas: Reporte que incluye, periodos de ingreso, año de ingreso, periodo de matrícula, años de matrículas, semestres, programa, apoyos, tipo de vulnerabilidad, datos de tipo etnográficos y domicilios, entre otros, 2015-1 a 2019-1
- Desertores: Reporte con datos de solo los estudiantes desertores por periodo, programa, promedio academico, semestre que cursaba al desertar, nombres y códigos estudiantiles desde 2015-1 al 2019-1
- Ayudas Educativas: Reporte con el tipo de ayudas educativas otorgadas a cada estudiante en cada periodo académico por programa del 2015 al 2019-1

- Remisiones a estrategias de atención: Reportes con datos de las estrategias en las que se atendieron los estudiantes, por profesional atendido, por programa, por fechas y por motivos.

Por otro lado, se solicitaron directamente al Departamento de Tecnología, la generación de los reportes relacionados con:

- Notas: Reporte de Calificaciones obtenidas por los estudiantes durante su tiempo de estudio en la Universidad en cada período académico, organizada por periodos académicos, cursos, grupos, profesores y calificación final obtenida 2014-1 al 2019-1
- Graduados: Reporte de estudiantes graduados, por programa, periodo y datos de identificación y nombre de estudiantes.
- Cursos: Reporte con código de los cursos por periodo por programa académico vigentes.
- Profesores: Reporte con nombres de profesores por programa por periodo.

Se consideraron y obtuvieron datos de los estudiantes, desde lo académico (como las notas), desde matriculas o ingreso (como estados, datos socioeconómicos y procedencia académica), individuales (desde la caracterización de la población desde la vulnerabilidad y tipo de apoyo para estudiar), los servicios ofrecidos y recibidos desde lo institucional para el (acompañamiento estudiantil, financiero y personal). Los datos se descargaron del sistema en archivos .csv, se validaron y estandarizaron.

5.3.2.2 CREACION DE BASE DE DATOS

De dichos ficheros se diseña la base de datos para el proyecto, con diferentes tablas en SQL, con información asociada, para posterior generación de datos correlacionados. El diseño de la base de datos inicial incluía la tablas de

1. Ingreso: con 35 campos asociados a la información de ingreso y datos socio-económicos y personales
2. Apoyos Financieros: cuenta con dos (2) campos que describen los tipos de apoyos existentes para los estudiantes
3. Deserción: cuenta con cinco (5) campos que describen las causas de deserción de estudiantes desertores y se asocia con campo a la tabla de ingreso.
4. Vulnerabilidad: dos (2) campos con la descripción del tipo de vulnerabilidad.
5. Catálogo de programas: dos (2) campos con los códigos y nombres de los programas académicos de pregrado
6. Profesor: con tres (3) campos con identificación y dato de contacto de profesores
7. Catálogo de cursos: con siete (7) campos con las descripciones de identificación de cursos y los periodos, año.
8. Grupo: Identificación del grupo asociado a cada curso con el profesor
9. Consejerías: con diez (10) campos que describen los tipos de consejerías y la información de la atención en fechas, periodos y cursos.
10. Calificaciones: con nueve (9) campos relacionados con la notas finales de cada curso, semestre del curso, programas y periodos.
11. Estrategias: con once (11) campos que describen las estrategias académicas en las cuales se les atendió a cada estudiante.

5.3.2.3 PREPROCESAMIENTO DE DATOS

En la fase de extracción de los datos, se tuvieron dificultades en la asociación de los cursos, grupos y profesores, ya que los datos se encontraban de manera incompleta en un alto porcentaje desde el origen, por lo tanto se decidió excluir de

la base de datos la tabla de profesor y grupo y se eliminaron de las tablas correlacionadas, el campo de asociación con las mismas, quedando finalmente un total de 11 tablas. Se define una vista única para almacenar los resultados extraídos de la base de datos, se empleó SQL (MS SQL para administrar los datos). Se diseñaron 4 consultas complejas que resultaron en los siguientes conjuntos de datos o vistas (Código disponible en Git Hub):

- **Matriculados con atención:** Se consolida toda la información relacionada con los estudiantes matriculados quienes registran atención de alguna de las estrategias y apoyos académicos que ofrece la universidad o externos. Se describe en la tabla 5.3.2.3:

MATRICULADOS CON ATENCIÓN: 16 ATRIBUTOS, 23362 REGISTROS	
APOYO	Describe el estado del estudiante para los apoyos económicos ofrecidos (SI, NO APLICA)
CONVENIO	Estado del estudiante en convenio con entidades externas de apoyo económico para estudiar (SI, NO APLICA)
CURSO	Nombre de los cursos matriculados en un período académico específico.
ESTADO	Estado de un estudiante en un año y periodo académico específico (Matriculado, Graduado o desertor)
ESTRATEGIA	Nombre de servicio de acompañamiento académico recibido (Monitoria, Tutoria Académica, Consejería académica)
ESTRATO	Clase socio económica al cual pertenece el estudiante (1,2, 3, 4, 5, 6)
GENERO	Género (Masculino, Femenino, Otro)
NOTAFINAL	Nota definitiva obtenida en un curso en un periodo de tiempo determinado (Rango entre 1,0 y 5,0)
NRODOC	Número de Documento de identificación oficial nacional del estudiante
NROMATRICULA	Código numérico asignado al estudiante para su identificación en la Universidad
PERIODO	Periodo académico en el cuál se tiene el registro del estudiante (1, 2)
PERIODOINGRESO	Periodo académico en el cuál el estudiante ingresó por primera vez a un programa de la universidad
PROGRAMA	Nombre de la carrera del estudiante
VULNERABLE	Estado de un estudiante en condición de vulnerabilidad (SI, NO APLICA)
YEAR	Año en el cuál se tiene el registro de datos del estudiante
YEARINGRESO	Año en el cuál ingresó el estudiante por primera vez

TABLA 5.1: Matriculados con atención.

- **Matriculados sin atención desde el 2018:** Estudiantes matriculados que no se les reporta ninguna atención en las estrategias y apoyos académicos que ofrece la universidad, se describen en la tabla 5.3.2.3:
- **Desertores con atención:** Estudiantes identificados por la universidad como desertores quienes en algún momento del periodo evaluado tuvieron alguna

MATRICULADOS SIN ATENCIÓN: 15 ATRIBUTOS, 559.114 REGISTROS	
NOMBRE ATRIBUTO	DESCRIPCIÓN
APOYO	Describe el estado del estudiante para los apoyos económicos ofrecidos (SI, NO APLICA)
CONVENIO	Estado del estudiante en convenio con entidades externas de apoyo económico para estudiar (SI, NO APLICA)
CURSO	Nombre de los cursos matriculados en un período académico específico.
ESTADO	Estado de un estudiante en un año y periodo académico específico (Matriculado, Graduado o desertor)
ESTRATO	Clase socio económica al cual pertenece el estudiante (1,2, 3, 4, 5, 6)
GENERO	Género (Masculino, Femenino, Otro)
NOTAFINAL	Nota definitiva obtenida en un curso en un periodo de tiempo determinado (Rango entre 1,0 y 5,0)
NRODOC	Número de Documento de identificación oficial nacional del estudiante
NROMATRICULA	Código numérico asignado al estudiante para su identificación en la Universidad
PERIODO	Periodo académico en el cuál se tiene el registro del estudiante (1, 2)
PERIODOINGRESO	Periodo académico en el cuál el estudiante ingresó por primera vez a un programa de la universidad
PROGRAMA	Nombre de la carrera del estudiante
VULNERABLE	Estado de un estudiante en condición de vulnerabilidad (SI, NO APLICA)
YEAR	Año en el cuál se tiene el registro de datos del estudiante
YEARINGRESO	Año en el cuál ingresó el estudiante por primera vez

TABLA 5.2: Matriculados sin atención.

atención en las estrategias y servicios de acompañamiento ofrecido por la universidad durante sus periodos matriculados. Se describe en la tabla 5.3.2.3:

DESERTORES CON ATENCIÓN: 18 ATRIBUTOS, 4214 REGISTROS	
NOMBRE ATRIBUTO	DESCRIPCIÓN
APOYO	Describe el estado del estudiante para los apoyos económicos ofrecidos (SI, NO APLICA)
CATEGORIA	Categoría de deserción del estudiante (ACADÉMICA, INDIVIDUAL, ECONÓMICO, INSTITUCIONAL)
CONVENIO	Estado del estudiante en convenio con entidades externas de apoyo económico para estudiar(SI, NO APLICA)
CURSO	Nombre de los cursos matriculados en un período académico específico.
ESTADO	Estado de un estudiante en un año y periodo académico específico (Matriculado, Graduado o desertor)
ESTRATEGIA	Nombre de servicio de acompañamiento académico recibido (Monitoria, Tutoria Académica, Consejería académica)
ESTRATO	Clase socioeconómica al cual pertenece el estudiante (1,2, 3, 4, 5, 6)
GENERO	Género (Masculino, Femenino, Otro)
NOTAFINAL	Nota definitiva obtenida en un curso en un periodo de tiempo determinado (Rango entre 1,0 y 5,0)
NRODOC	Número de Documento de identificación oficial nacional del estudiante
NROMATRICULA	Código numérico asignado al estudiante para su identificación en la Universidad
PERIODO	Periodo académico en el cuál se tiene el registro del estudiante (1, 2)
PERIODOINGRESO	Periodo académico en el cuál el estudiante ingresó por primera vez a un programa de la universidad
PROGRAMA	Nombre de la carrera del estudiante
SEMESTRE	Semestre en el cual estaba matriculado el estudiante en el periodo y año específico (1 a 10)
VULNERABLE	Estado de un estudiante en condición de vulnerabilidad (SI, NO APLICA)
YEAR	Año en el cuál se tiene el registro de datos del estudiante
YEARINGRESO	Año en el cuál ingresó el estudiante por primera vez

TABLA 5.3: Desertores con atención

- Desertores sin atención: Aquellos estudiantes identificados como desertores por la universidad sin ninguna atención registrada durante sus periodos de matrícula. Se describe en la tabla 5.3.2.3:

DESERTORES SIN ATENCION: 16 ATRIBUTOS. 77159 REGISTROS	
NOMBRE ATRIBUTO	DESCRIPCIÓN
NRODOC	Número de Documento de identificación oficial nacional del estudiante
NROMATRICULA	Código numérico asignado al estudiante para su identificación en la Universidad
CONVENIO	Estado del estudiante en convenio con entidades externas de apoyo económico para estudiar (SI, NO APLICA)
APOYO	Describe el estado del estudiante para los apoyos económicos ofrecidos (SI, NO APLICA)
VULNERABLE	Estado de un estudiante en condición de vulnerabilidad (SI, NO APLICA)
ESTADO	Estado de un estudiante en un año y periodo académico específico (Matriculado, Graduado o desertor)
ESTRATO	Clase socioeconómica al cual pertenece el estudiante (1,2, 3, 4, 5, 6)
GENERO	Género (Masculino, Femenino, Otro)
YEARINGRESO	Año en el cuál ingresó el estudiante por primera vez
PERIODOINGRESO	Periodo académico en el cuál el estudiante ingresó por primera vez a un programa de la universidad
CATEGORIA	Categoría de deserción del estudiante (ACADEMICA, INDIVIDUAL, ECONOMICO, INSTITUCIONAL)
PROGRAMA	Nombre de la carrera del estudiante
CURSO	Nombre de los cursos matriculados en un período académico específico.
NOTAFINAL	Nota definitiva obtenida en un curso en un periodo de tiempo determinado (Rango entre 1,0 y 5,0)
YEAR	Año en el cuál se tiene el registro de datos del estudiante
PERIODO	Periodo académico en el cuál se tiene el registro del estudiante (1, 2)

TABLA 5.4: Desertores sin atención.

Estas 4 vistas se exportaron a archivos .csv. Posteriormente se pasó a validar, estandarizar los datos de los 4 archivos, se descartaron los registros con datos incompletos procedimiento que se realizó con lenguaje R en R Studio. (Código disponible en Git Hub).

5.3.3 ANÁLISIS EXPLORATORIO DE DATOS

Luego de la familiarización de los datos y la revisión exploratoria, encontramos que las causas de deserción mas incidentes (obtenidas mediante encuestas y cuestionarios de salida) se agruparon en cuatro (4) categorías como se muestra en 5.3.3: Como se puede observar, los principales factores o dimensiones fueron los económi-

Académico	Económico	Individual	Institucional
13.7 %	40.9 %	45 %	0.4 %

TABLA 5.5: Porcentajes de Categorías incidentes en la deserción

cos en especial la falta de financiamiento o baja solvencia, la dimensión individual elementos personales, familiares y de salud, luego sigue el académico por variables relacionadas con pérdida y repitencia y por último menos incidente pero igual relevante, el institucional donde se agruparon las causas que resultaron imputables a la institución.

También se agruparon los estudiantes, en la tabla 5.3.3 según el estrato económico, el grueso de los estudiantes se encuentran en estratos 1, 2 y 3:

1	2	3	4	5	6
32 %	35 %	24 %	7 %	1 %	0.6 %

TABLA 5.6: Distribución de porcentaje de estudiantes por estratos

Y se obtuvo la distribución de las categorías de deserción por estratos como lo muestra la tabla 5.3.3:

Factor / Estrato	1	2	3	4	5	6
Académico	3.7 %	4.4 %	4.2 %	1.2 %	0.1 %	0.1 %
Económico	13.2 %	15.2 %	9.4 %	2.5 %	0.2 %	0.3 %
Individual	15.1 %	15.3 %	10.4 %	3.2 %	0.9 %	0.2 %
Institucional	0.1 %	0.2 %	0.1 %	0.1 %	0 %	0 %

TABLA 5.7: Porcentaje de estudiantes por Categorías de deserción y Estratos

Se señala que al relacionar los cuatro (4) factores o dimensiones con los estratos, la mayor parte de desertores se centran en los estratos 1, 2 y 3 y en las dimensiones económicos e individuales como se había anticipado. En los estratos 4, 5, 6 tiene mayor relevancia la parte individual aunque también se ve la incidencia del económico. La tasa de deserción acumulada por programas por cohortes en el periodo (contra ingresos en el periodo) 5.3.3 se distribuyó de la siguiente manera:

5.3.3.1 REGRESIÓN MULTINOMIAL

Se aplicó la regresión multinomial para los estudiantes caracterizados como desertores y que recibieron apoyo académico. Se seleccionaron las variables:

- APOYO (Factor w/ 35 levels)
- Vulnerable (Factor w/ 8 levels)
- Estado (Factor w/ 3 levels)
- Estrato (int)
- Genero (Factor w/ 2 levels)
- YearIngreso (int)

Programa	Tasa (%)
Ingeniería de Mercados	66
Microbiología	55
Ingeniería de Sistemas	55
Contaduría Pública	50
Ingeniería Multimedia	44
Comercio y Negocios Internacionales	36
Medicina	32
Trabajo Social	30
Psicología	27
Derecho	26
Administración de Empresas	22
Ingeniería Industrial	22
Fisioterapia	19
Instrumentación Quirúrgica	17
Enfermería	14

TABLA 5.8: Tasa de deserción por programas académicos.

- PeriodoIngreso (int)
- PROGRAMA (Factor w/ 15 levels)
- CURSO (Factor w/ 731 levels)
- NotaFinal (num)
- Year (int)
- PERIODO (int)
- IDCausa (int)
- categoria (Factor w/ 4 levels)

Se aplicó un algoritmo de regresión multinomial para identificar los factores de mayor incidencia en las categorías de deserción establecidas (variable Categoría).

Para dividir los conjuntos de entrenamiento y prueba se empleó un valor de corte de 0.8.

Se balancearon y estandarizaron los datos de entrada, se empleo un valor confiabilidad de 0.95 y se utilizó una validación cruzada de 5fold.

Finalmente se obtuvo un valor de asertividad de 77% en la clasificación de nuevos casos.

Esta asertividad se puede mejorar, agregando elementos, que se planean tener en cuenta en futuros modelos.

A partir de los resultados se observó que las variables más significativas para identificar los desertores fueron .Apoyos "Vulnerabilidad" la pérdida de apoyos por parte de la universidad desde lo ofrecido en la excelencia académica así como los ofertados desde Bienestar institucional y la pérdida de financiamiento de las entidades nacionales de apoyo como el ICETEX y desde lo individual, se relaciona con el estado y condición de vulnerabilidad del estudiante, relacionado con la poca adaptación a la vida universitaria. (Código disponible en Git Hub).

Si bien esta primera caracterización arrojó una congruencia entre las categorías de deserción y el historial de los estudiantes la generalización de la misma a otros estudiantes es cuestionable dado que la muestra es muy pequeña (solo estudiantes matriculados a partir de 2018 de estos un muy pequeño porcentaje, menor del 5 % de los matricualdos en ese periodo. Se puede concluir con los resultados obtenidos, que el estrato social influye, en los primeros semestres es mas alto el porcentaje de deserción, las notas de reprobado influirían en la deserción, la tendencia de deserción en mayor proporción para aquellos estudiantes quienes se matriculan en el primer período del año, el programa de derecho y psicología tienen mayor número de estudiantes y mayor deserción.

Se realizaron otras estadísticas para lograr realizar un mejor análisis de los atributos los cuales se pueden visualizar en el anexo **Anexo 3**.

5.3.4 CONSTRUCCIÓN DE LAS ESCALAS DE PREFERENCIA DE LOS ATRIBUTOS.

Desde el resultado del análisis exploratorio se crea un único fichero de datos, consolidando todos los campos necesarios para la caracterización de los estudiantes, resultando un fichero de 21 campos y 635.540 registros, como se puede observar en la tabla siguiente 5.3.4:

DATOS CONSOLIDADOS: 21 CAMPOS 635.540 REGISTROS	
NOMBRE DEL CAMPO	DESCRIPCIÓN
APOYO	Describe el estado del estudiante para los apoyos económicos ofrecidos (SI, NO APLICA)
CALIFICACIONES	Escala de medición de notas obtenidas clasificadas en (EXCELENTE, BUENO, APROBADO Y REPROBADO)
CATEGORIA	Categoría de deserción del estudiante (ACADEMICA, INDIVIDUAL, ECONOMICO, INSTITUCIONAL)
COHORTE	Describe el año y período de ingreso de los estudiantes por primera vez
CONVENIO	Estado del estudiante en convenio con entidades externas de apoyo económico para estudiar (SI, NO APLICA)
CURSO	Nombre de los cursos matriculados en un período académico específico.
ESTADO	Estado de un estudiante en un año y período académico específico (Matriculado, Graduado o desertor)
ESTRATEGIA	Nombre de servicio de acompañamiento académico recibido (Monitoria, Tutoria Académica, Consejería académica)
ESTRATO	Clase socio económica al cual pertenece el estudiante (1, 2, 3, 4, 5, 6)
GENERO	Género (Masculino, Femenino, Otro)
NOTAFINAL	Nota definitiva obtenida en un curso en un período de tiempo determinado (Rango entre 1,0 y 5,0)
NRODOC	Número de Documento de identificación oficial nacional del estudiante
PERIODO	Período académico en el cual se tiene el registro del estudiante (1, 2)
PERIODOINGRESO	Período académico en el cual el estudiante ingresó por primera vez a un programa de la universidad
PERIODOS.DESDE.INGRESO	Períodos transcurridos desde el primer ingreso
PROGRAMA	Nombre de la carrera del estudiante
SEMESTRE	Estado de un estudiante en condición de vulnerabilidad (SI, NO APLICA)
VULNERABLE	Año en el cual se tiene el registro de datos del estudiante
YEAR	Año en el cual ingresó el estudiante por primera vez
YEARINGRESO	Año en el cual ingresó el estudiante por primera vez
YEARS.DESDE.ULTIMA.MATRICULA	Nro de años que han pasado desde la última matrícula

TABLA 5.9: Fichero consolidado

5.3.4.1 CREACIÓN DE CONJUNTO CONSOLIDADO DE DATOS

Los requisitos de datos se establecen considerando factores académicos, económicos, individuales e institucionales que se extraerán del sistema ERP en USB. En este caso se dispuso de los datos disponibles al momento de realizar este trabajo los siguientes indicadores, entre otros no considerados relevantes para este trabajo, para cada alumno en cada año y semestre de alta como alumno en la USB:

- Académico: Documento Oficial de Identidad (NroDoc), un código de identidad único asignado dentro de la Universidad (NroMatricula), Primer año (Year) y Semestre (S) en el que se inscribió en USB, cada año y semestre que se inscribió como estudiante en USB con registro del profesorado, programa y cursos cursados (al menos 1 curso y como máximo 7 cursos), así como las calificaciones obtenidas en cada curso, finalmente también se asignó una categoría académica a cada alumno: matriculado, graduado y egresado, si no se tiene registro en un semestre también se registró una causa principal de deserción (YR, SR, Facultad, Programa, Curso, Calificaciones, Estado, Categoría de deserción).
- Económicas: ayudas económicas concedidas al alumno en un año y semestre (FA *). El estatus socioeconómico, en Colombia se mide como “Estratos” definido en una escala del 1 al 6, donde 6 son los más ricos y 1 los más pobres, (NSE *).
- Individual: si es considerado Vulnerable (posee alguna discapacidad o cualquier otra condición reconocida que lo haga vulnerable) (V *).
- Institucional: si participa en un programa de acompañamiento institucional (consejería psicológica, académica o financiera) (CP *).

Las características marcadas con (*) pueden tener más de un valor por alumno.

5.3.4.2 ASIGNACIÓN DE ETIQUETAS DE "DESERTOR" O "NO DESERTOR" A CADA ALTERNATIVA.

Se definen los procedimientos (códigos) para identificar a los estudiantes que no se matriculan en al menos dos periodos consecutivos para etiquetarlos como Desertores, de lo contrario se identifican como "No Desertores".

Se añade un nuevo campo que establece las coincidencias o diferencias de esta caracterización respecto a la caracterización previa realizada por la institución.

5.3.4.3 CARACTERIZACIÓN DE DESERTORES Y REDUCCIÓN DE ATRIBUTOS.

Desde el fichero final obtenido con 21 atributos y luego de la dedicación exhaustiva para la depuración de los datos y la transformación de los mismos (90 % del tiempo), se agregaron dos campos calculados para representar las categorías de deserción: en un campo los estudiantes fueron clasificados por cada año (YEAR) y el semestre (Semestre) ha sido abandonado (SÍ) o no (NO) y en otro campo calculado los estudiantes fueron clasificados como dados de baja siguiendo la definición dada en Colombia: "desertor por discontinuidad es un estudiante que no se registra en la institución por un período que comprende más de un semestre contiguo", o "no se ha graduado entre el intervalo de tiempo de 10 por semestres más desde su primer registro en la institución ", desertor por Cohorte. En cualquier otro caso, el estudiante está etiquetado como (NO DESERTOR).

Para las dos nuevas columnas se calculó si un estudiante ha tenido una calificación de NULL en todos los cursos ofrecidos en un YEAR-SEMESTRE se considera que abandonó, y si un estudiante lo abandona en más de un semestre consecutivo, entonces se etiqueta como DESERTOR para todos esos semestres, si en diez o más semestres un estudiante ha no graduado, entonces se etiquetó como DESERTOR

POR COHORTE, de lo contrario, se etiquetó como NODESERTOR (no desertor).

Los campos con muchos valores deben transformarse como discretos con pocos valores, como, por ejemplo, las calificaciones que en Colombia pueden tomar valores entre 1 y 5 (las calificaciones se asignaron a Reprobado, Aprobado, Bueno, Excelente).

Se reasignaron algunos atributos como, por ejemplo, Apoyo y Estrategia que se transformaron a una escala de SÍ o NO, indicando si un estudiante recibió ayuda económica o participó en un programa de acompañamiento (SÍ) o no (NO).

Como el volumen de información disponible es grande y complejo, ML se utilizará para realizar la reducción de características y simplificar la cantidad de datos que se utilizarán como entrada para el problema de decisión. Como no hay evidencia de qué método funciona mejor, se emplearán los métodos más populares y sus resultados se compararán para extraer las características más sólidas. Entonces se consideran las siguientes variables para el análisis así:

Variables	Descripción
Calificaciones	Calificaciones obtenidas (Excelente, Bueno, aceptable,)
Years desde última matrícula	Nro de años que han pasado desde la última matrícula
Estado	Estado (Matriculado, Graduado, Desertor)
Year	Año relacionado con cursos semestrales registrados
Estrategia	Intervención por parte de programa de acompañamiento, si o no aplica
Apoyo	Ayudas externas financieras, Si o No aplica
Género	Género
Facultad	Facultad
Convenio	Ayuda para financiamiento de la Universidad, Si o No aplica
Vulnerable	Tipo de Vulnerabilidad, si aplica
Estrato	Estrato o clase socio económica
Periodo	Semestre, relacionado con el año y registro de cursos
IS Desertor	Desertor por cohorte, desertor por semestre, No desertor
Desertor Si No	Desertor o No desertor (matriculado o graduado)

La función Automodel de Rapidminer se aplicó al conjunto de datos construido. Se aplicaron los siguientes métodos para predecir tanto Desertor SiNo como Is Desertor:

- Naive Bayes
- Generalized Linear Model
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine

Se obtiene el automodel:

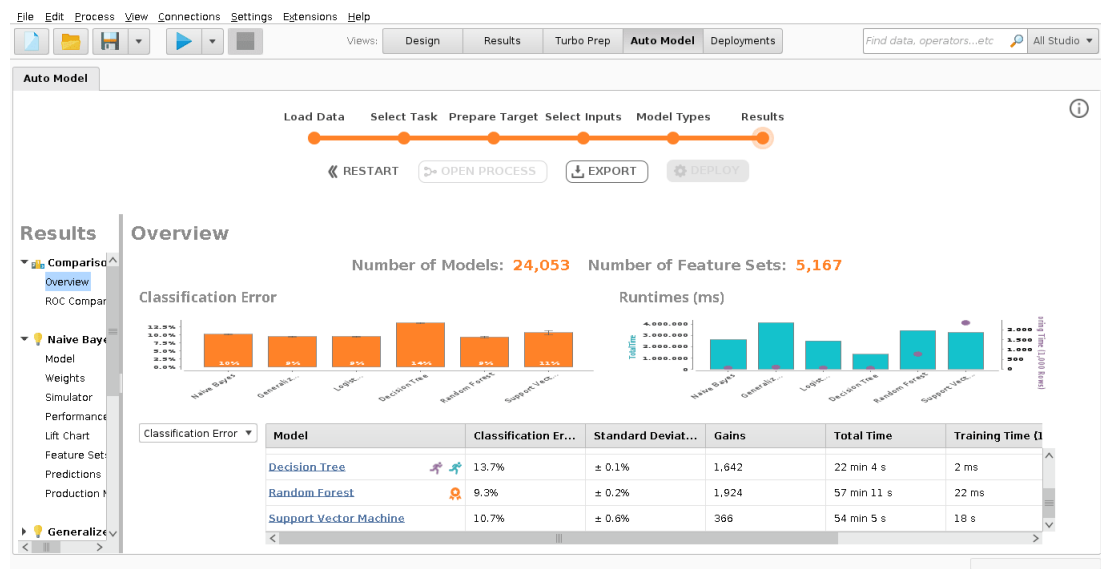


FIGURA 5.9: Comparativo Resultados Modelos

La precisión de la clasificación se muestra en la Tabla 5.3.4.3 (con 70/30 por ciento de datos divididos para entrenamiento y pruebas). Los números entre paréntesis representan la clasificación del método por precisión. Como se puede ver en la Tabla 5.3.4.3, el generalizado se comportó mejor en ambas situaciones, seguido de

la regresión logística, pero en general todos los métodos funcionaron bien. Además, los peores algoritmos fueron Naïve Bayes y Support Vector Machines (aunque no se empleó kernel).

	DESERTOR SI_NO	IS_DESERTOR
Method	Accuracy	
Naive Bayes	90.00 % (4)	89.93 % (4)
Generalized Linear Model	90.20 % (1)	90.57 % (2)
Logistic Regression	90.11 % (3)	90.52 % (3)
Decision Tree	90.16 % (2)	86.31 % (5)
Random Forest	89.97 % (5)	90.70 % (1)
Support Vector Machine	87.70 % (6)	89.33 % (6)

TABLA 5.10: Comparativo resultados de Modelos

Se obtuvieron resultados similares con otras métricas de calidad como AUC, especificidad, Recall, sensibilidad, F-Measure. Vale la pena señalar que las clases generadas por ambas columnas están altamente desequilibradas, en métricas generales como, por ejemplo, AUC, precisión, Sensibilidad, no dan resultados precisos. Por lo tanto, se emplearon otras métricas como G-MEAN y Balanced Accuracy, pero se obtuvieron resultados similares

Al examinar en detalle la clasificación de la columna **IsDesertor**, se puede observar que todos los algoritmos se desempeñaron mal en la clase de identificación "DESERTOR POR PERIODO", este resultado se puede explicar por tres hechos: hay muy pocos casos para esta clase y no hay patrón regular reconocible para estos estudiantes.

Como resultado de la selección de características, las características más seleccionadas en ambas situaciones por todos los métodos fueron: Calificaciones, YEARS.desde.la.última.matrícula, Estado, Convenio, YEAR, Genero. Perteneciente a categorías o dimensiones Académica, Económica e Individual. Los mismos que fueron los más relevantes

para la deserción de los estudiantes según las respuestas dadas por un cuestionario que se les hizo a los estudiantes. Como se muestra en la Figura 5.10

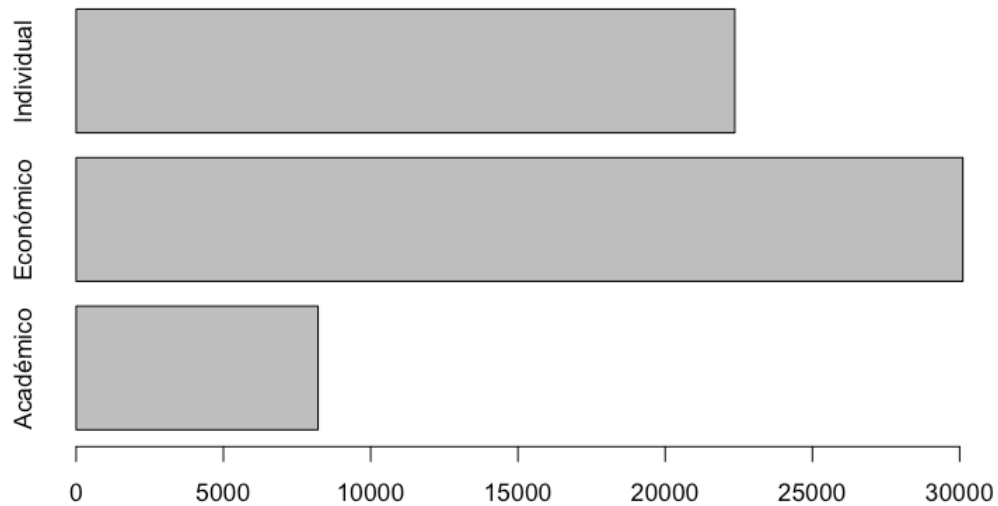


FIGURA 5.10: Principales razones de deserción

Al aplicar en Rapidminer automodel solo para Modelo Lineal Generalizado y Regresión Logística para los atributos seleccionados para clasificar ambos atributos (DesertorSiNo y IsDesertor), los resultados de precisión se pueden observar en la tabla 5.3.4.3:

	DESERTOR SI_NO	IS_DESERTOR
Method	Accuracy	
Generalized Linear Model	90.16 % (1)	90.50 % (2)
Logistic Regression	90.09 % (2)	90.60 % (1)

Se observa que la regresión logística se comporta ligeramente mejor que el modelo lineal generalizado (GML) al clasificar la columna ISDesertor.

Se aplicó un criterio de mayoría simple para la jerarquización a partir de la suma de votos (1) si fueron seleccionados por un método y (0) si no fueron seleccionados. De igual manera, se identifican por colores los mejores en cada método (verde), los segundos mejores (Naranja) y los terceros (lila). Igualmente los de mayor sumatoria de los escogidos son los Amarillos y los azul cielo los segundos más escogidos.

Para el caso de DESERTORSINO tenemos:

ATRIBUTO / MODELO	DESERTORSINO						SUMATORIA
	NAIVES BAYES	GENERALIZED LINEAR MODEL	LOGISTIC REGRESSION	DECISION TREE	RANDOM FOREST	SUPPORT VECTOR MACHINE	
APOYO	1	1	0	0	0	0	2
CALIFICACIONES	1	1	1	0	1	1	5
CONVENIO	0	1	0	0	1	1	3
ESTADO	1	1	1	1	1	1	6
ESTRATEGIA	0	1	0	0	0	0	1
ESTRATO	0	1	0	0	0	0	1
FACULTAD	0	1	0	0	0	0	1
GENERO	0	1	1	0	0	0	2
PERIODO	0	1	0	0	0	1	2
PERIODOINGRESO	0	1	0	0	0	0	1
VULNERABLE	0	1	0	0	0	0	1
YEAR	0	1	1	0	0	1	3
YEARS.DESDE.ULTIMA.MATRICULA	1	1	1	0	1	0	4

TABLA 5.11: Selección de features por mayoría Simple

Se observa que los mejores en cada método son : Calificaciones y Estado, los segundos mejores Convenio, Estado y Years.desde.ultima.matricula y terceros mejores por métodos, estado, Género, Vulnerable, Year y Year desde ultima matrícula.

Igualmente los de mayor sumatoria de los escogidos son Calificaciones, Estado y Year desde ultima matricula y los segundos mas escogidos Convenio y YEAR.

Para el caso de ISDESERTOR tenemos:

ATRIBUTO / MODELO	ISDESECTOR						SUMATORIA
	NAIVES BAYES	GENERALIZED LINEAR MODEL	LOGISTIC REGRESSION	DECISION TREE	RANDOM FOREST	SUPPORT VECTOR MACHINE	
ESTADO	1	1	1	1	1	0	5
YEARS.DESDE.ULTIMA.MATRICULA	1	1	1	1	1	1	6
CALIFICACIONES	1	1	1	1	1	0	5
APOYO	0	1	0	0	0	0	1
CONVENIO	0	1	1	1		0	3
ESTADO	0	1	0	0	0	0	1
ESTRATEGIA	0	1	1	0	0	0	2
ESTRATO	0	1	0	0	0	0	1
GENERO	0	1	1	1		0	3
PERIODO	0	1	0	0	0	0	1
VULNERABLE	0	1	0	0	0	0	1
YEAR	0	1	1	0	0	0	2
FACULTAD	0	1	1	0	0	0	2
ESTADO for DESERTOR POR PERIODOS	0	0	0	0	0	1	1
ESTRATO for DESERTOR POR PERIODOS	0	0	0	0	0	1	1

TABLA 5.12: Selección de features por mayoría Simple

Se observa que los mejores en cada método son : Calificaciones, Estrategia y Years.desde.ultima.matricula. Segundos mejores por método son Estado, Years desde ultima matricula y Estrato for Desertorporperiodos. Terceros mejores por métodos: calificaciones, estado, estado por desertorporperiodo, years desde ultima matricula.

Igualmente los de mayor sumatoria de los escogidos son Estado,Year desde ultima matricula, Calificaciones y los segundos mas escogidos Convenio y Género.

5.3.5 FORMULACIÓN DE LA RECOMENDACIÓN.(PASO 3)

Esta parte no se considera en la tesis, por el tiempo que requiere así como la disposición de recursos por parte de la institución y que cumpla con la implantación de las observaciones derivadas de la aplicación de la fase de diagnóstico.

CAPÍTULO 6

CONCLUSIONES

En este capítulo se plantean las conclusiones y trabajo futuro.

6.0.1 CONCLUSIONES

Se estudia el problema de la permanencia estudiantil en instituciones de educación superior desde una perspectiva multidimensional considerando las dimensiones reconocidas en la literatura publicada pero incorporando variables propuestas por la autora. El resultado se recoge en una metodología de gestión de la permanencia que se presenta en este trabajo, en el capítulo 4.

También en el capítulo 4 se incluye una metodología de apoyo a la decisión para apoyar la toma de decisiones para la gestión eficiente de la permanencia, esta metodología se basa en dos metodologías generales: Crip-DM y MCDM, integrando la ciencia de datos con la toma de decisiones bajo criterios múltiples lo cual constituye un aporte interesante tanto para la gestión de la permanencia como la para ciencia de datos y la toma de decisiones bajo múltiples criterios.

La componente MCDM de la metodología aporta estructura al proceso, mientras que el componente relativo a la ciencia de datos de la metodología ayuda a hacer frente a la naturaleza masiva y diversa de los datos que deben tenerse en cuenta pa-

ra el desarrollo de estrategias y políticas para abordar eficazmente la deserción de estudiantes en la educación superior.

Se aplica la metodología a un caso real de estudio, la Universidad Simón Bolívar sede Barranquilla, Colombia, en el que se muestra como llevar a cabo el proceso complejo de diagnosticar y luego elaborar los elementos para tomar las decisiones en cuanto a la generación de políticas y estrategias para la gestión de la permanencia.

Resulta interesante el uso de Rapidminer para la caracterización de la deserción empleando la herramienta de Automodel que resultó ser muy eficiente al aplicarse a un caso complejo que se aborda por volumen de datos, el fuerte desbalanceo de clases y la cantidad de diferentes variables que se consideran. Demostró ser una herramienta confiable en general, si bien se requirió de un hardware potente para poder llegar a los resultados en unos pocos días.

Desde el resultado del análisis, se facilita la realización de planes de mejora en función a lo esperado del modelo de permanencia implementado en la institución, y de esta manera fortalecer la capacidad institucional en pro a la mejora de la permanencia y graduación estudiantil, minimizando la deserción.

Los resultados de la aplicación de la metodología al estudio de caso han demostrado que los estudiantes que eventualmente abandonan son muy difíciles de clasificar porque no exhiben patrones reconocibles, al menos dentro del alcance de los datos recopilados en Universidad Simón Bolívar.

6.0.2 TRABAJO FUTURO

Se considera trabajar en la profundización del conocimiento de las características de los estudiantes lo mas detallado y estándar posible, para poder lograr identificar patrones reconocibles faltantes y generar las estrategias y acciones efectivas para el grupo poblacional con el objetivo de avanzar en el nivel de madurez del modelo

de gestión de la permanencia implementado en la Universidad. De igual manera, proponer la metodología a nivel nacional, en instituciones de educación de vecindad, para iniciar un observatorio de la permanencia y graduación estudiantil en cuanto a los modelos de gestión implementados.

Desde el punto de vista metodológico hay que perfeccionar los aspectos de elicitación de preferencias de los decisores, pues en general son personas muy ocupadas que no tienen tiempo disponible y realmente son ellos quienes deben exponer su subjetividad para que sea modelada, en este sentido sería un área interesante de investigación generar los modelos de preferencias con apoyo de técnicas de IA que reduzca el esfuerzo del decisor sin sacrificar la exactitud y veracidad del modelo.

Por último se recomienda desarrollar un sistema de apoyo a la decisión que implemente la metodología para facilitar su adopción e implementación.

APÉNDICE A

ANEXO 1

A.1 TABLA DE EVALUACIÓN

Criterios	Actividad	Indicadores	SIN	IMPACTO	IMPACTO	IMPACTO
			IMPACTO	MÍNIMO	SUFICIENTE	IMPORTANTE
Gestión y autoevaluación	Proceso de admisión	Identificación de posibles desertores Recomendación de acompañamiento para posibles desertores Integrada en el sistema de calidad				
	Gestión de la permanencia del estudiante	Integrada en el sistema de enseñanza-aprendizaje Evaluación o fortalecimiento de capacidades de estudiantes				
Seguimiento y apoyo académico	Evaluación / Fortalecimiento de capacidades de estudiantes	Se aplican exámenes de diagnóstico Fomento y apoyo a grupos de estudio Capacitación tutores				
	Mecanismos de asesoría y tutoría	Disponibilidad de espacios y recursos para tutorías y asesorías Integración con otras actividades académicas				
Apoyos Económicos	Asesoría, otorgamiento y seguimiento de apoyos	Existen estrategias, políticas para otorgar apoyos				
		Existen mecanismos para asignación de apoyos externos a estudiantes				
		Existen mecanismo de seguimiento a los apoyos asignados a los estudiantes.				
	Evaluación de impacto de apoyos en permanencia y calidad de resultados académicos	Evaluación del impacto en la permanencia de los apoyos económicos otorgados. Los resultados de la evaluación del impacto son accesibles y están integrados con el sistema de enseñanza-aprendizaje.				
Seguimiento de la situación individual y personal	Seguimiento y actualización de situación	Existen mecanismos simples y efectivos para actualizar la situación individual o personal del estudiante cuando se ocupe				
		Existen mecanismos para dar seguimiento a la situación individual o personal del estudiante cuando se ocupe				
Cultura de la información	Gestión y autoevaluación	Existen estrategias para la automatización de la gestión de la información relacionada con la gestión de la Permanencia.				
		Existen políticas (reglamentos) para la publicación, acceso y protección de la información.				
	Seguimiento y apoyo académico	Está implementada la hoja de vida del estudiante.				
		Generación de reportes de seguimiento en tiempo real.				
	Apoyos económicos	Acceso y disponibilidad a la información.				
		Mecanismos simples y flexibles para actualizar la información.				
	Situación individual o personal	Esta automatizada la gestión de la información relacionada con apoyos económicos.				
		Está integrado al sistema de enseñanza-aprendizaje la información sobre apoyos económicos.				
		Existe un sistema, módulo para gestionar, actualizar información relacionada con la situación individual y/o personal del estudiante.				
		La gestión de la información del seguimiento está integrada con otros sistemas informáticos de la institución.				

APÉNDICE B

ANEXO 2

B.1 ORIGEN DE DATOS CASO DE ESTUDIO

En la Universidad Simón Bolívar, se venía utilizando un sistema de información académica y financiera denominada SIAFF, la cual tenía toda la información de estudiantes y administrativos, de manera general. Desde el año 2017 se inició el proceso de revisión frente al seguimiento efectivo a los estudiantes de la universidad Simón Bolívar en pregrado que apoyara de alguna manera la aplicación de actividades y estrategias que permitieran mejorar la calidad académica de los estudiantes desde la detección de las variables incidentes en los estudiantes, con el apoyo de toda la comunidad académica involucrada y de esta manera inicia la construcción de módulos en el sistema de información que permitieran remitir, informar y revisar resultados de manera sencilla para el análisis del comportamiento de los estudiantes, buscando tener la información relacionada con las estrategias académicas y/o psicológicas que recibían los estudiantes en el sistema de información, en el cual se encontraban las notas de los estudiantes, los tipos de matriculas e informacion socio económica principalmente. En el año 2011 se implementó en el Sistema Institucional de Información Académico y Financiero (SIAF) para llevar el control de la deserción en la Universidad Simón Bolívar, desde entonces se ha ido perfeccionando el sistema y se han ido agregando los datos pertenecientes a ambos periodos durante 10 años.

En este trabajo se incluyen los resultados obtenidos al aplicar la metodología general de la ciencia de datos propuesta por IBM al conjunto (iV), dónde se utilizaron datos de desertores que no recibieron apoyos o atención institucional por diversas causas.

Aunque se disponía de datos de deserción e ingreso desde el 2015, es desde el 2018 que se estandariza la información que se recoge en el sistema, es por ello que se decidió emplear datos desde ese año para el análisis.

En una primera fase, se definen los módulos relacionados con la recogida de información precisa y adecuada de las estrategias de acompañamiento académico, denominado tutorías, dónde se registra la información por cursos, por profesor, por programa, por unidad de apoyo académico y por fecha en la cuál se presta el servicio y a cuál estudiante se le presta. De igual manera con el módulo de la estrategia de consejería académica, servicio académico / Psicológico, el cual se ofrece desde la coordinación del Programa Institucional de Excelencia Académica (PIEA), en la cual a través de un formulario en la web diligencian los campos relacionados con la atención de cada estudiante, por programa, psicólogo, curso y fecha. Posteriormente se realiza el módulo de la estrategia de Monitoria académica, el cual registra información de Monitores (Estudiantes destacados por programas y unidades académicas), los cuales desarrollan las actividades de acompañamiento a sus pares compañeros quienes requieren refuerzo de temas en cursos con mayor dificultad y se registran en el sistema a través de un formulario en la web, la atención, por cursos, fechas y programa académico. Paralelamente se desarrollan informes dónde se identifican cursos críticos para consulta por parte de los programas académicos, con toda la información relacionada con los profesores, grupos y porcentajes de estudiantes quienes pierden los cursos en cada corte parcial evaluativo definido por la Universidad. Así como se afina el reporte de pérdidas de cada corte parcial, con el fin de identificar estudiantes críticos en cada corte parcial e iniciar una intervención directa. Luego, se realiza el módulo de remisión a estudiantes, en el cual, según Arrieta Matos (2020), se involucra directamente a los profesores en el proceso de seguimiento a los estudiantes dentro del aula, identificando en ellos, algunas posibles variables que puedan

incidir en la deserción o la baja calidad académica. En conclusión: Se requiere la información de matriculas de los estudiantes, tipos de ayudas, condición de vulnerabilidad, notas, programa, periodos de deserción, caracterización socio económica e individual en general.

APÉNDICE C

ANEXO 3

C.1 RESULTADOS DE ESTADISTICAS EN R

Entre las estadísticas generadas con el fichero consolidado en r tenemos:

RESULTADOS:

Ayudas: Se realiza un primer comparativo de los estudiantes desertores y no desertores y las ayudas económicas recibidas durante el tiempo matriculado, obteniendo la siguiente relación como muestra la figura C.1:

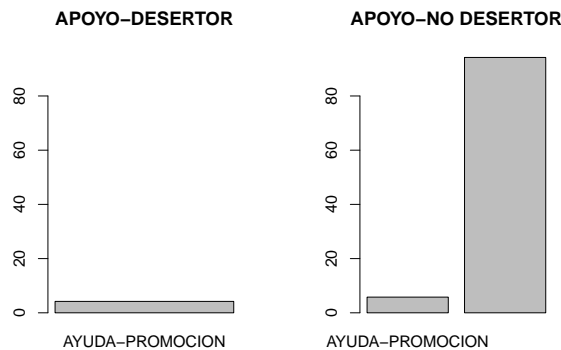


FIGURA C.1: Comparativo Desertor no desertor y ayudas

En la comparación podemos ver el apoyo económico a pesar de ser menos en

desertores, tiene un comportamiento muy similar con los no desertores, lo que nos puede indicar que no es lo suficientemente relevante la variable de ayudas educativas.

Calificaciones: En esta segunda comparación se tienen en cuenta las calificaciones de los estudiantes en las categorías de desertores y no desertores, desde la clasificación de Aprobado y reprobado como muestra la figura C.2:

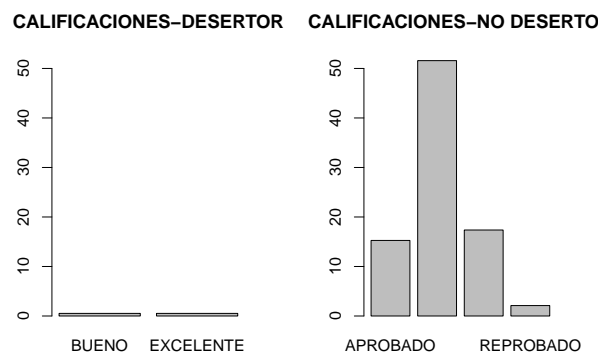


FIGURA C.2: Comparativo Calificaciones: Excelente, Bueno, Aprobado, Reprobado

Podemos observar que la barra de reprobados en desertores es mayor que en la de los no desertores.

Categorías de deserción: Se comparan los desertores y no desertores con las categorías de Deserción así como muestra la figura C.3:

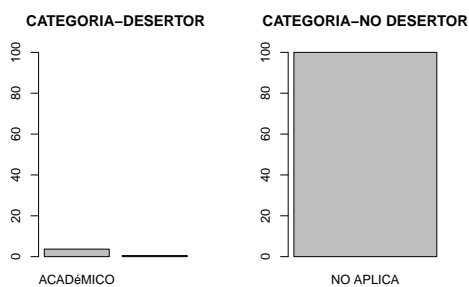


FIGURA C.3: Comparativo Categorías de deserción: Académico, Institucional, Económico, Individual(Antropológico)

Los no desertores no aplica la categoria de desertores. En el caso de desertores, encontramos que se tienen en cuenta las categorías académico (2 primeras barras), económico (2 siguientes barras), individual (antropológico), institucional y NA. Podemos ver que las categorías mas altas son: Individual y Económica.

Cohortes de desertores por Programas: En esta comparación se comparan todos los programas académicos de la universidad y su comportamiento en cuanto a cuales son los programas con mayor deserción y de cual cohorte así como muestra la figura C.4:

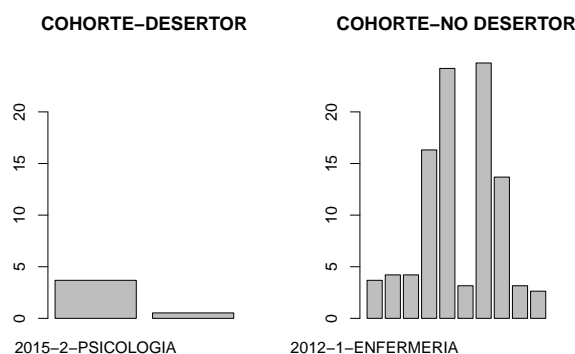


FIGURA C.4: Comparativo Cohortes Desertores Programas

Se puede observar que para el caso de las cohortes de desertores, se ve un pico en el programa de Psicología en la cohorte del año 2015-2.

Convenios: Se realiza el comparativo de los desertores y no desertores con los convenios que pudieron tener cada uno de los estudiantes durante su paso por la universidad así como muestra la figura C.5:

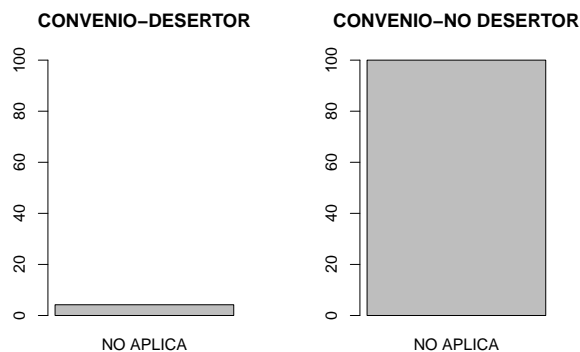


FIGURA C.5: Comparativo Convenios

Se observa en la comparación de estudiantes desertores con convenio de ayudas educativas regionales, se evidencia que estudiantes desertores contaron con la ayuda educativa.

Estados: Se definen los diferentes estados que puede tener un estudiante (Desertor, Matriculado y graduado) y se comparan con los desertores y no desertores así como muestra la figura C.6:

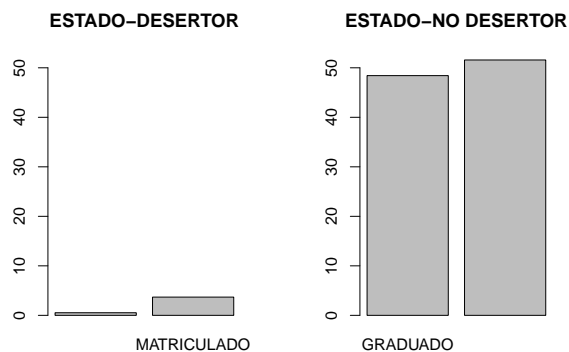


FIGURA C.6: Comparativo Estados

Los estados de los estudiantes desertores, pasan en su gran mayoría a ser estudiantes matriculados, y en algún momento a ser graduados. Dentro de la clasificación de los matriculados o no desertores, se tiene una proporción mínima de estudiantes

con estados de desertor.

Estrategia: Se realiza comparativo de las categorías de desertor y no desertor, con las estrategias de acompañamiento académico ofrecidas en la universidad así como muestra la figura C.7:

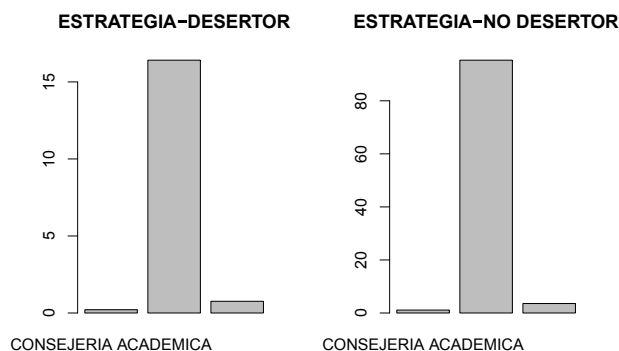


FIGURA C.7: Comparativo Estrategia

Los estudiantes quienes recibieron tutoria y consejería académica entre los desertores, han sido pocos. Sin embargo, el mas alto porcentaje de desertores, no recibieron ninguna estrategia. Aun en lo no desertores, el porcentaje de quienes han recibido apoyos es bajo.

Estrato: Se realiza comparativo de los desertores y no desertores, con los estratos socio económicos establecidos en Colombia en una escala de 1 a 6, siendo 1 el más bajo y 6 el más alto así como muestra la figura C.8:

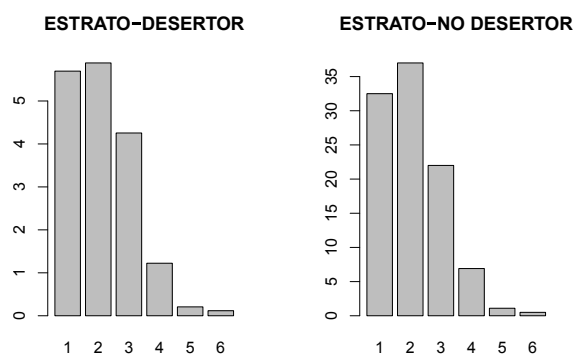


FIGURA C.8: Comparativo Estrato

Se puede observar que los estudiantes desertores se concentran en su mayoría

en los estratos mas bajos, 1, 2 y 3. Al igual que en los no desertores, la población de la universidad se concentra en los tres primeros estratos.

Género: Se realiza comparativo con el género (masculino y fememino) definidos en la escala así como muestra la figura C.9:

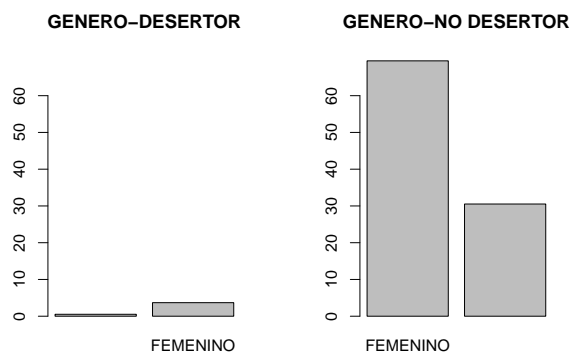


FIGURA C.9: Comparativo Género

Se conoce que la mayor población de la universidad es de género Femenino y de acuerdo a la comparación realizada, tiene un número mayor de desertores, así como un mayor número de matriculados.

Desertores por período: Se comparan los desertores y no desertores, con los estudiantes desertores por periodo académico 1 (enero-Julio), y 2 (agosto-Diciembre) (ausencia de matricula en dos periodos académicos semestrales consecutivos) así como muestra la figura C.10:

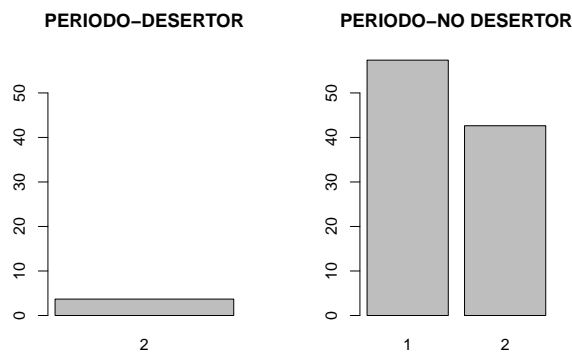


FIGURA C.10: Comparativo Desertores por periodo

El porcentaje de desertores por periodo indica que en el segundo periodo se da mayor deserción.

Periodo de Ingreso: Se comparan los periodos de ingreso semestral 1 (enero-Julio), y 2 (agosto-Diciembre) con los desertores y no desertores así como muestra la figura C.11:

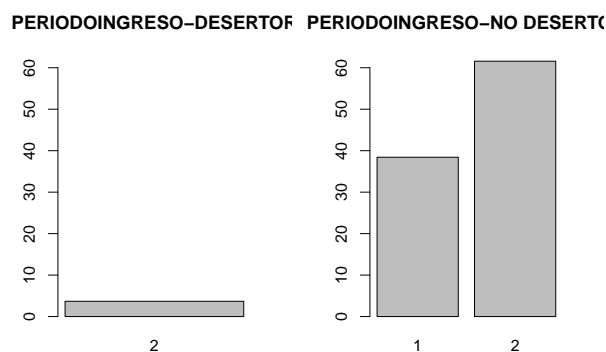


FIGURA C.11: Comparativo Desertores por periodo de ingreso

Se puede observar que el mayor porcentaje de desertores se evidencia en aquellos quienes ingresaron en el segundo semestre del año.

Programas: Se realiza comparativo de nro de desertores y no desertores con los programas académicos así como muestra la figura C.12:

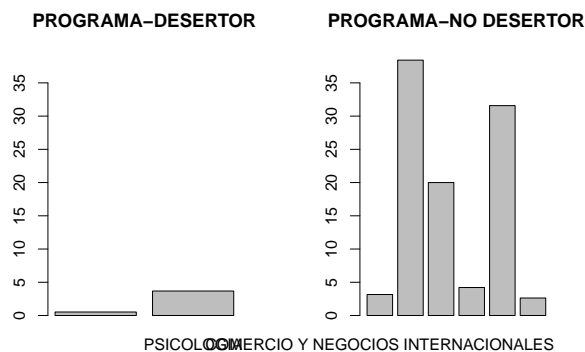


FIGURA C.12: Comparativo Programas y desertores

Los programas que mas deserción presentan son Psicología y comercio y negocios internacionales.

Vulnerabilidad: Se realiza comparativo de acuerdo a criterio de vulnerabilidad asociados a los desertores y no desertores en la caracterización, como muestra la figura C.13:

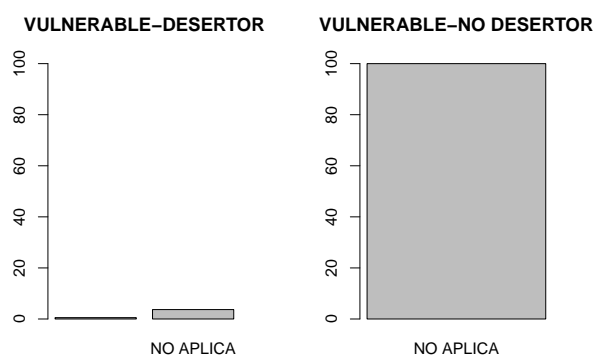


FIGURA C.13: Comparativo vulnerabilidad

El resultado indica que la vulnerabilidad o discapacidad incide en bajo porcentaje en la deserción. Sin embargo, hay estudiantes vulnerables con deserción por periodo. La barra mas alta indica No Aplica.

Años de deserción: Se compara los años dónde se evidencia la deserción con

los desertores y no desertores, como muestra la figura C.14:

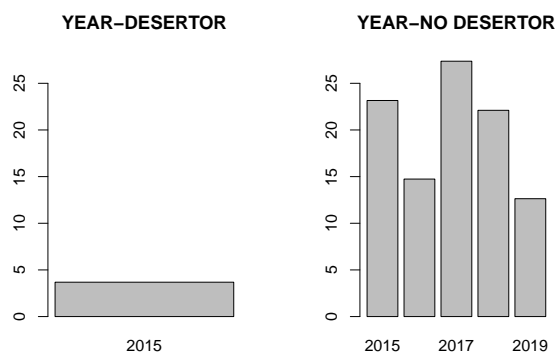


FIGURA C.14: Comparativo Desertores por año

Se puede observar que en el registro, el número de desertores ha disminuido al pasar los años, evidenciando un pico en 2015.

Año de ingreso del desertor: Se compara el año de ingreso registrado del desertor y no desertor, como muestra la figura C.15:

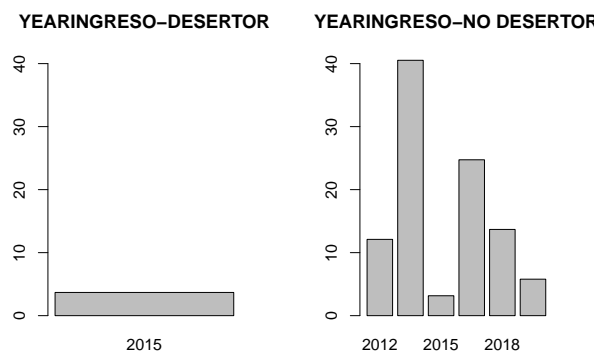


FIGURA C.15: Comparativo Año de ingreso vs Desertor

Se evidencia que en el año 2015 el pico de deserción es mas alto que los demás años.

Años de última matrícula: Se hace comparativo de la cantidad de años que han pasado desde la última matrícula con el desertor y no desertor, como muestra

la figura C.16:

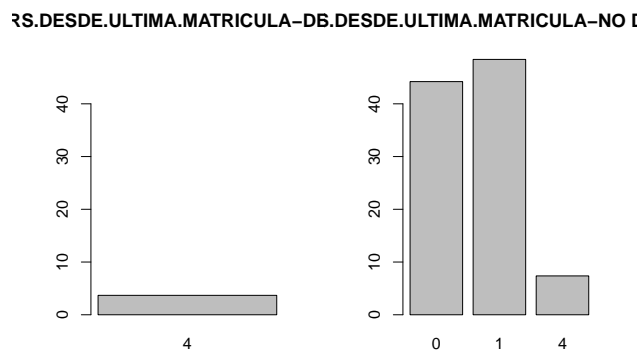


FIGURA C.16: Comparativo Años desde la última matrícula

La gráfica nos muestra la cantidad de años que han pasado desde la ultima matricula a corte del período 20191, mostrando que los desertores con mayor porcentaje son los de año 4.

Graduados y desertores: Comparativo de estudiantes graduados con desertores y no desertores, como muestra la figura C.17:

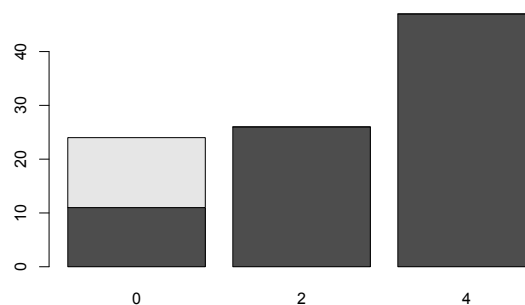


FIGURA C.17: Comparativo Graduados y desertores

La figura muestra el comparativo de los estudiantes graduados, no graduados y desertores durante un periodo de tiempo. La zona gris indica a los graduados.

Cohortes y Deserción: Se realiza un comparativo de las cohortes y el nro de años de abandono o no graduados, como muestra la figura C.18:

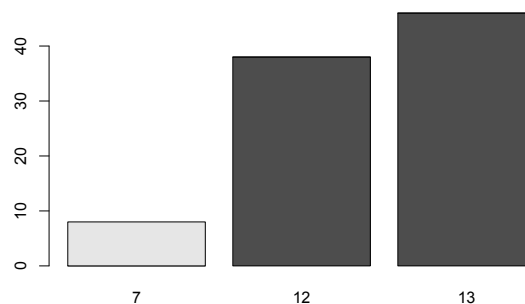


FIGURA C.18: Comparativo Cohortes Desertores Programas

La gráfica muestra que después de ciertos periodos académicos los estudiantes abandonan. El número de años esta entre 7 y 15 años.

APÉNDICE D

ANEXO 4

D.1 ESTADÍSTICAS DE RAPID MINER

La función Automodel de Rapidminer se aplicó al conjunto de datos construido. Se aplicaron los siguientes métodos para predecir tanto Desertor Si No como Is Desertor:

- Naive Bayes
- Generalized Linear Model
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine

Dentro de las estadísticas generadas para cada uno de los procesos realizados al considerar el análisis desde cada Atributo así:

DESERTOR SINO

Apoyos:

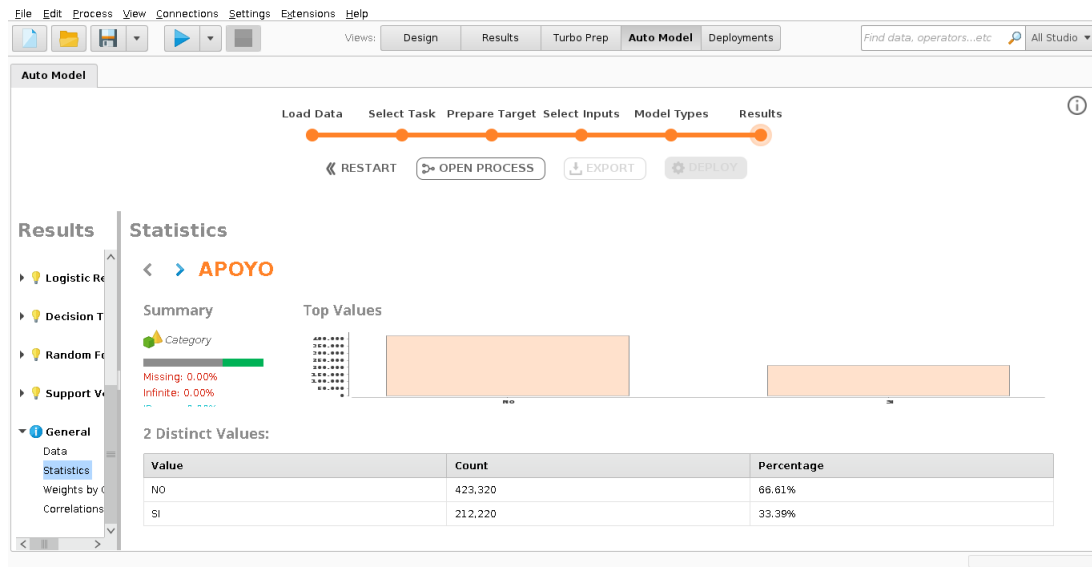


FIGURA D.1: Apoyo

Calificaciones:

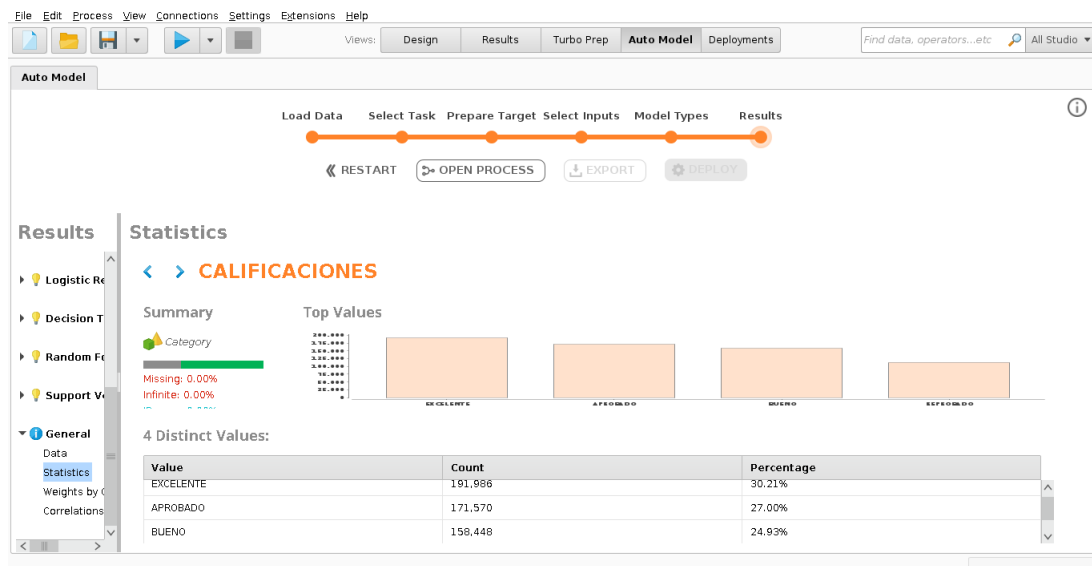


FIGURA D.2: Calificaciones

Convenios:

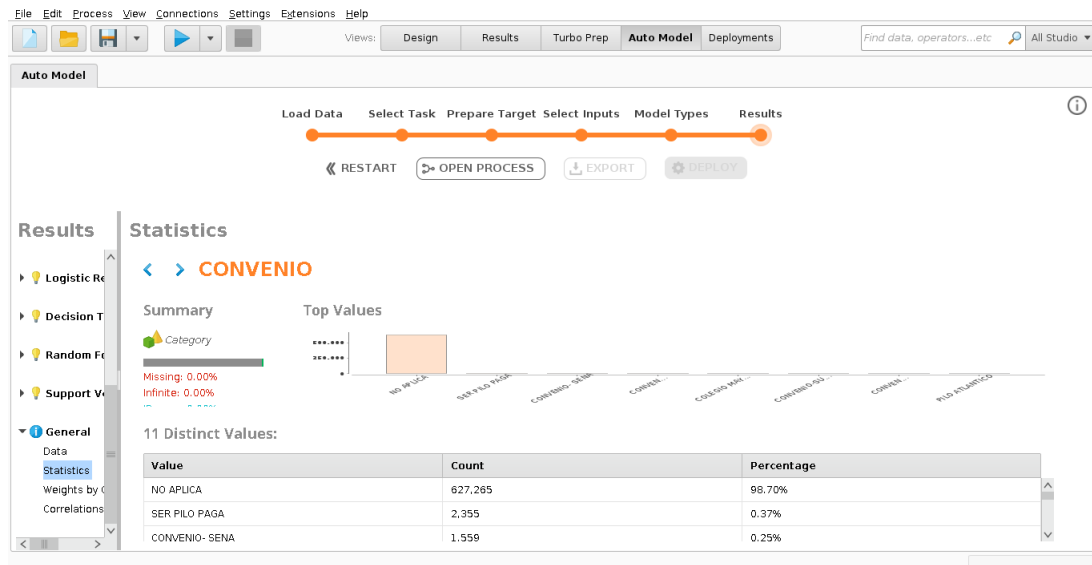


FIGURA D.3: Convenios

Estados:

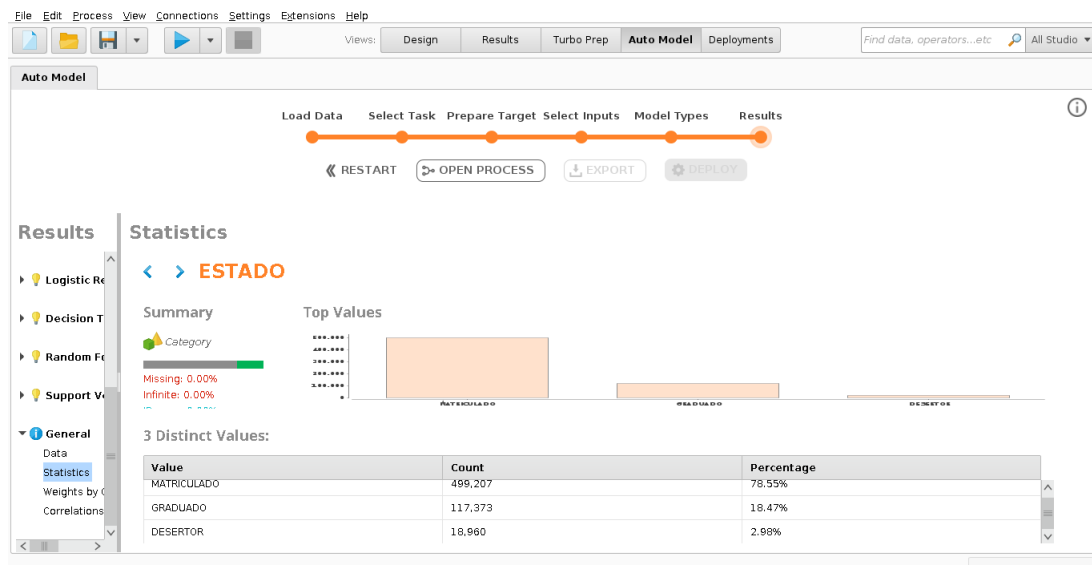


FIGURA D.4: Estados

Estrategias:

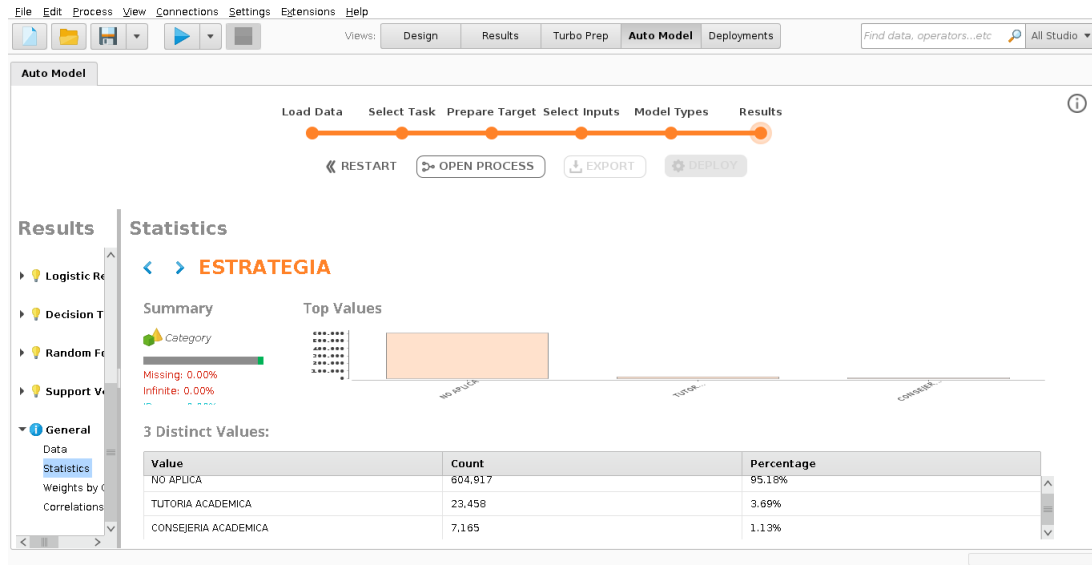


FIGURA D.5: Estrategias

Estratos:

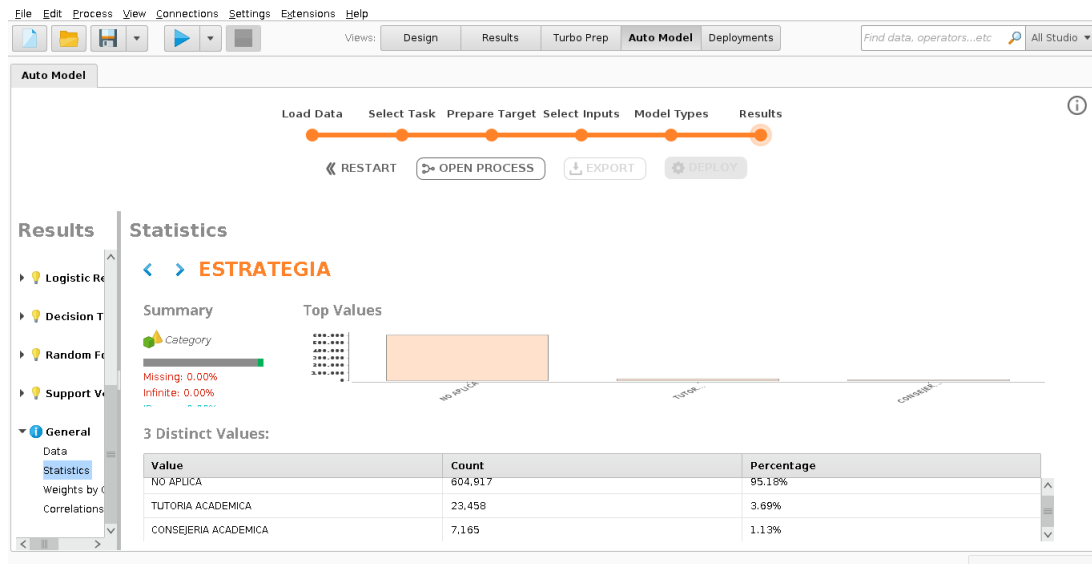


FIGURA D.6: Estratos

Facultad

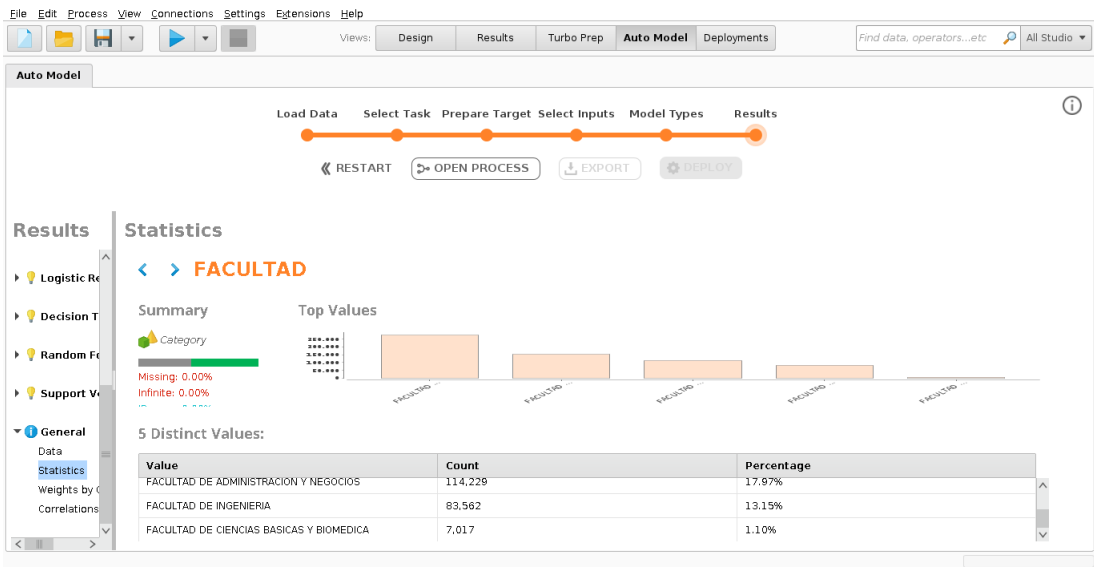


FIGURA D.7: Facultades

Se tiene un consolidado de los pesos obtenidos para cada atributo en cada método así:

NAIVES BAYES	
ATRIBUTO	Weights
CALIFICACIONES	0,209
YEARS.DESDE.ULTIMA.MATRICULA	0,198
ESTADO	0,069
APOYO	0,048
GENERALIZED LINEAR MODEL	
ATRIBUTO	Weights
CALIFICACIONES	0,22
YEARS.DESDE.ULTIMA.MATRICULA	0,177
VULNERABLE	0,064
ESTADO	0,053
ESTRATEGIA	0,053
PERIODOINGRESO	0,04
APOYO	0,039
PERIODO	0,038
ESTRATO	0,035
YEAR	0,028
FACULTAD	0,025
GENERO	0,021
CONVENIO	0,004

LOGISTIC REGRESSION	
ATRIBUTO	Weights
CALIFICACIONES	0,227
YEARS.DESDE.ULTIMA.MATRICULA	0,187
GENERO	0,09
ESTADO	0,063
YEAR	0,06
DECISION TREE	
ATRIBUTO	Weights
ESTADO	0,224
RANDOM FOREST	
ATRIBUTO	Weights
CALIFICACIONES	0,23
ESTADO	0,216
YEARS.DESDE.ULTIMA.MATRICULA	0,036
CONVENIO	0
SUPPORT VECTOR MACHINE	
ATRIBUTO	Weights
ESTADO	0,175
CALIFICACIONES	0,173
YEAR	0,066
CONVENIO	0,021
PERIODO	0,019

ISDESERTOR

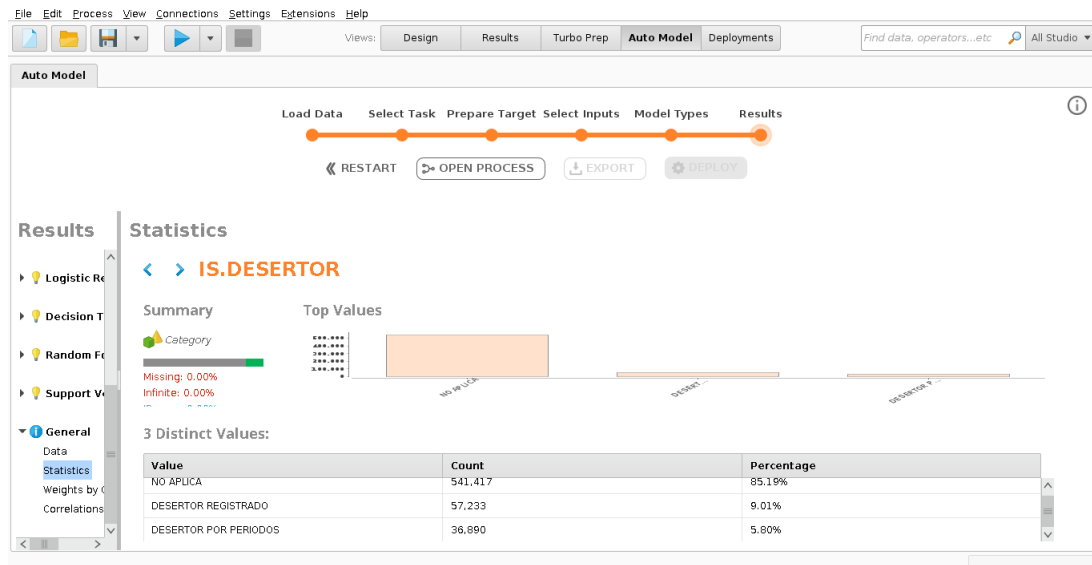


FIGURA D.8: Isdesertor

Periodo

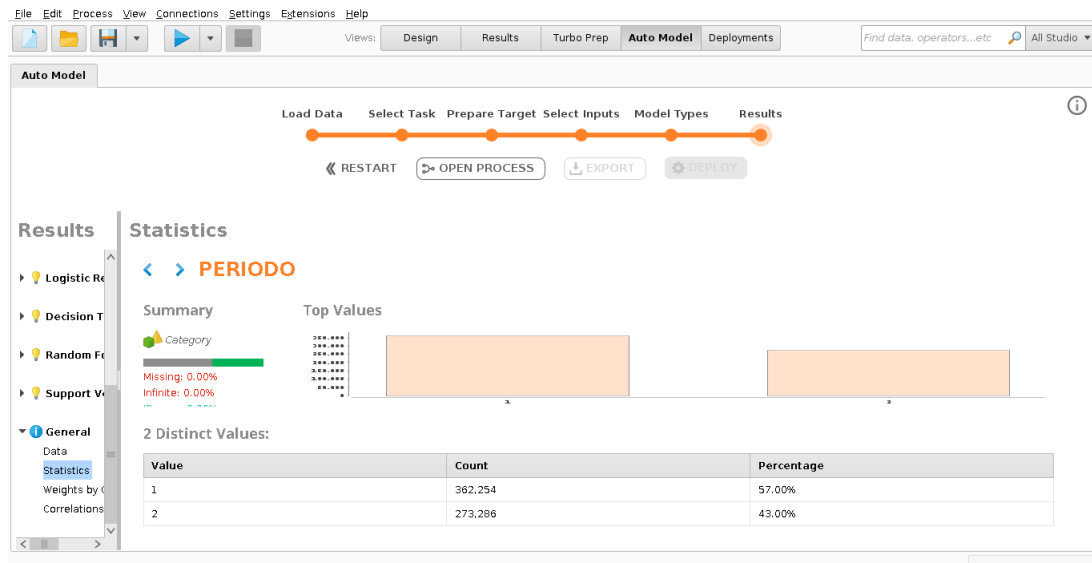


FIGURA D.9: Períodos

Vulnerable

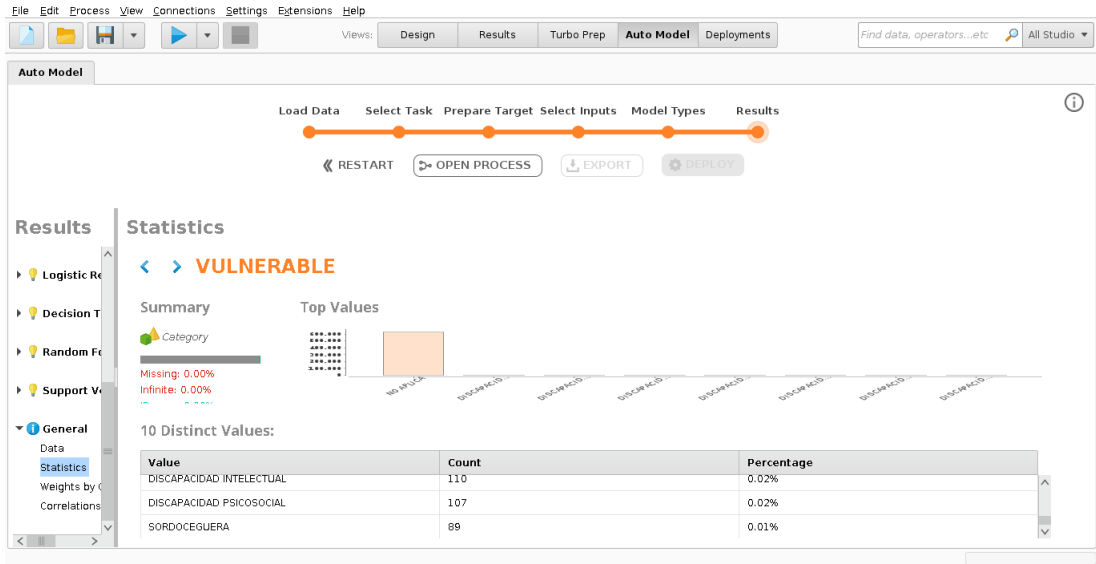


FIGURA D.10: Vulnerable

Año

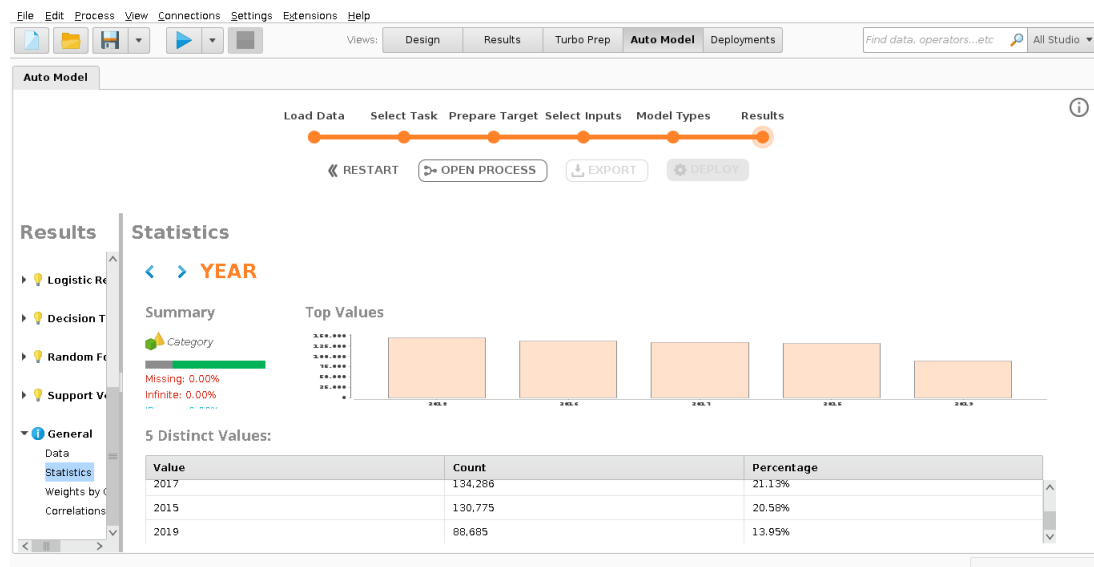


FIGURA D.11: YEAR

Años de la última matrícula

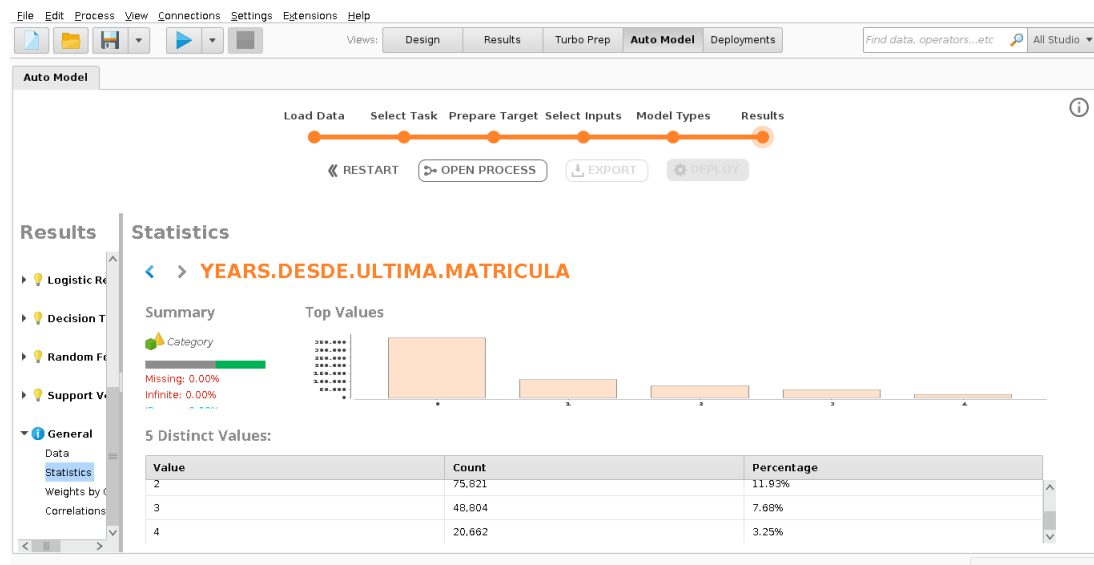


FIGURA D.12: Year.Desde.Última.Matricula

ISDESERTOR: Para el caso se obtuvieron las siguientes estadísticas:

Calificaciones

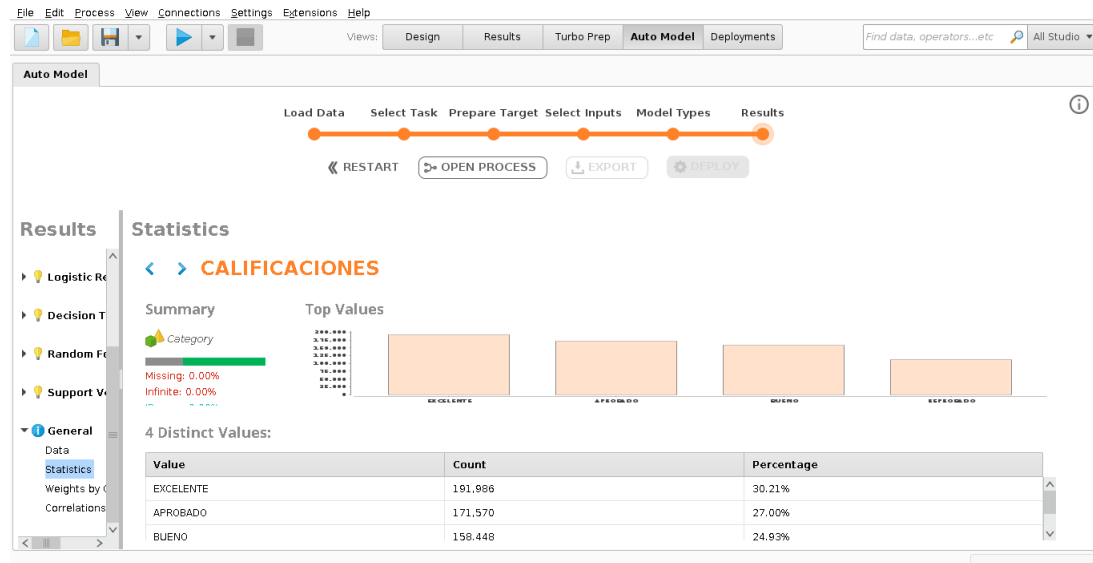


FIGURA D.13: Calificaciones

Desertor SINO

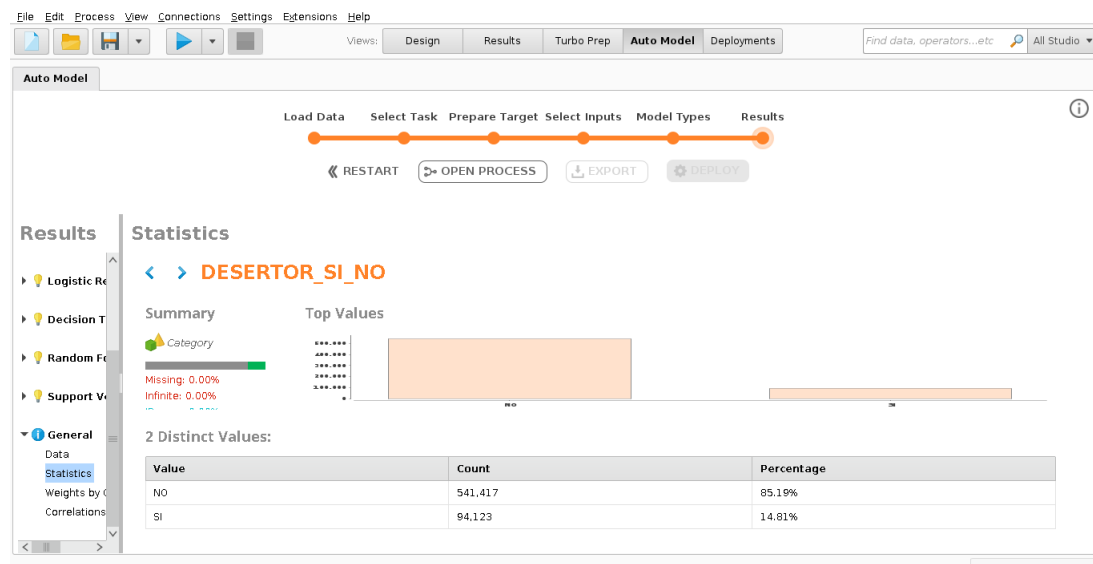


FIGURA D.14: DesertorSINO

Estado

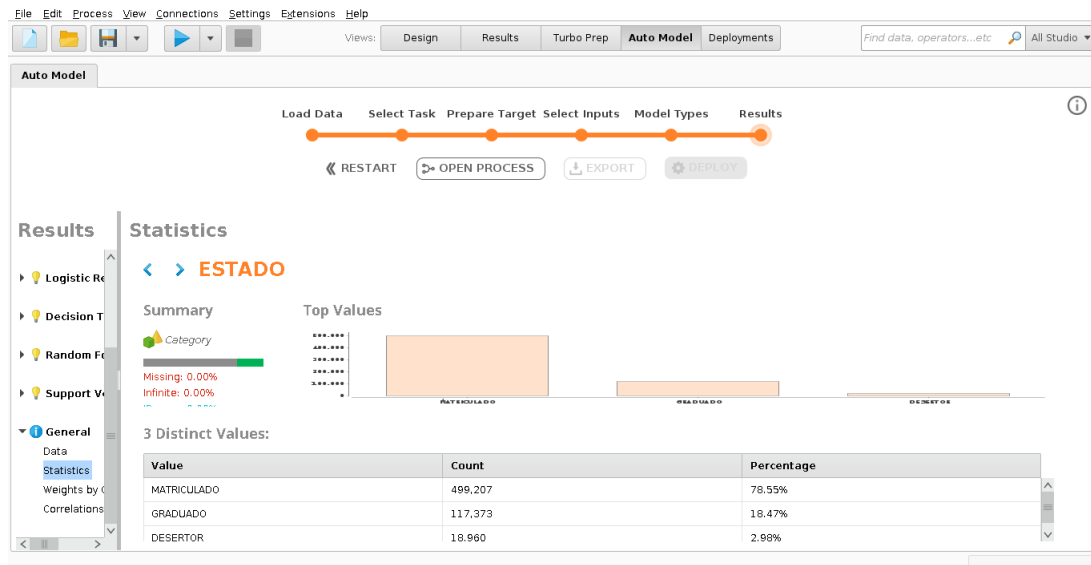


FIGURA D.15: Estado

Estrategias

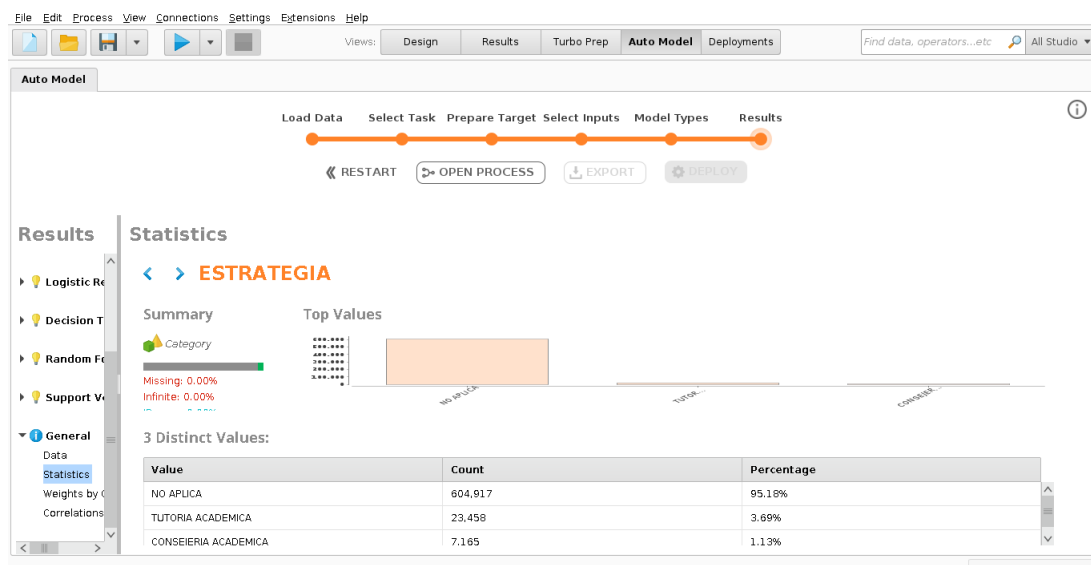


FIGURA D.16: Estrategias

Facultad

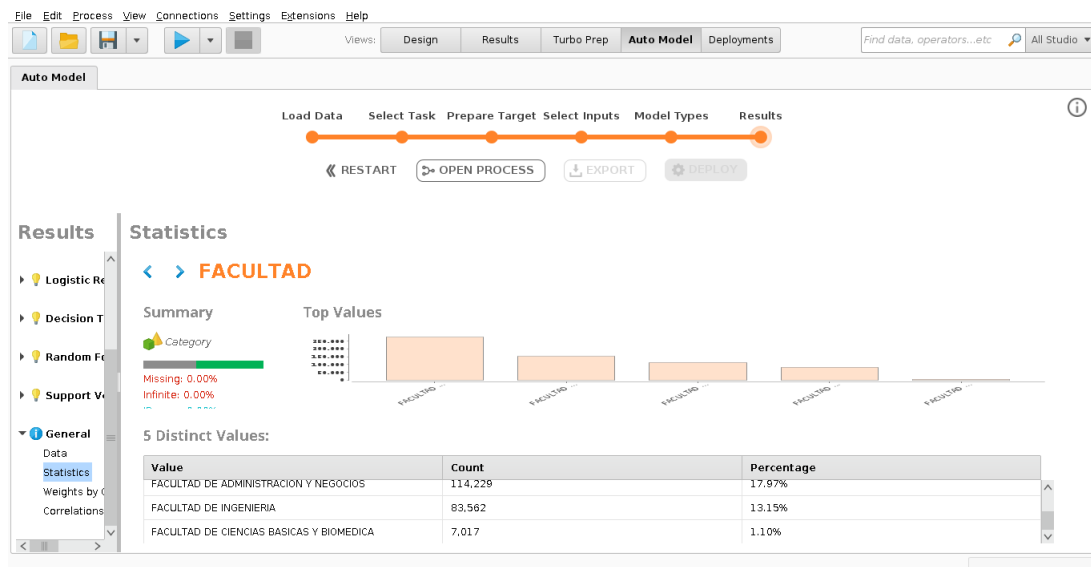


FIGURA D.17: Facultad

Genero

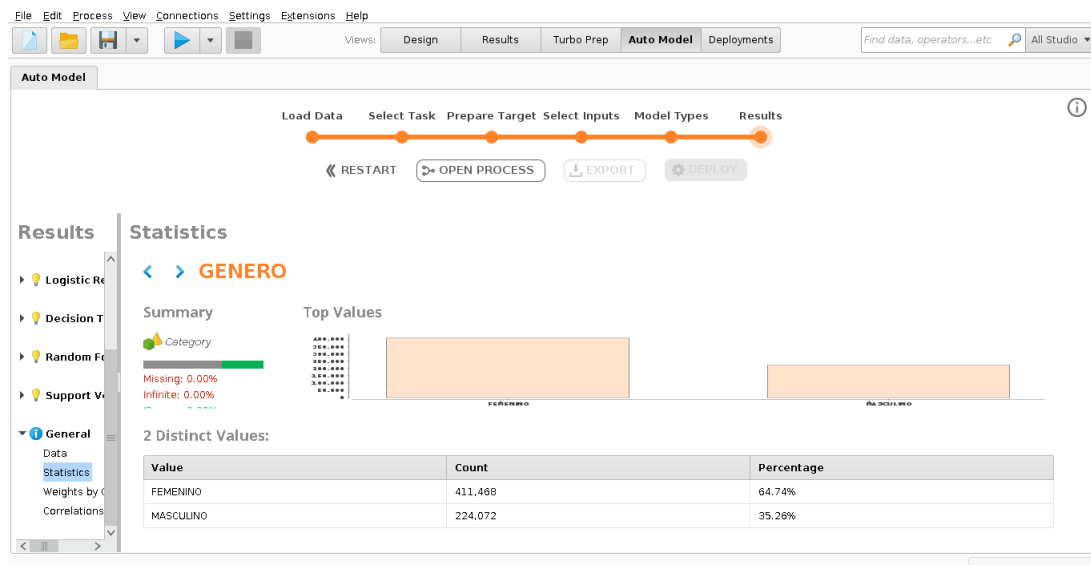


FIGURA D.18: Género

PERÍODO

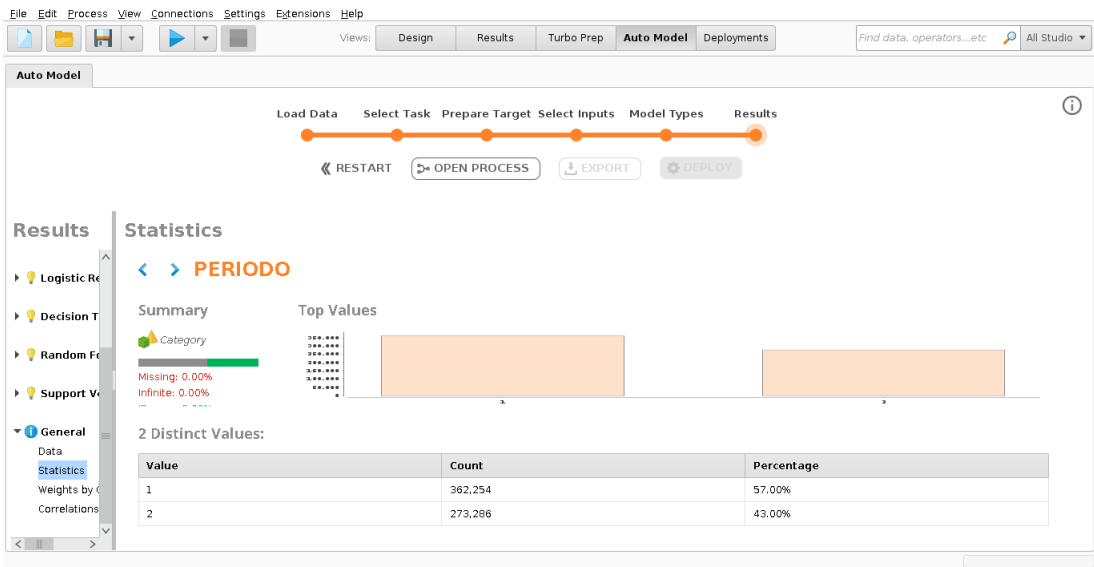


FIGURA D.19: Período

PERÍODO DE INGRESO

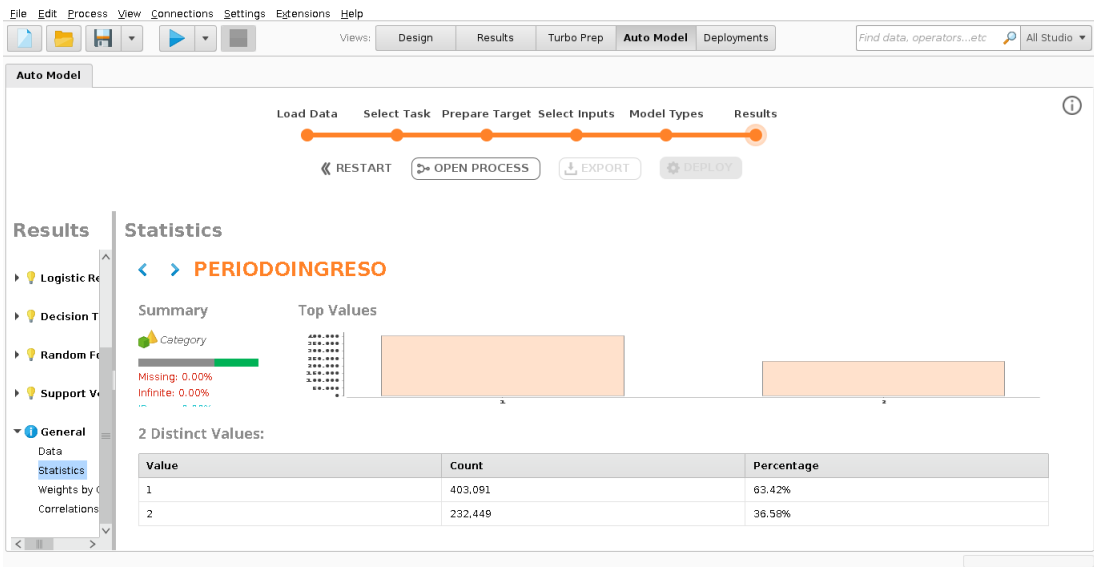


FIGURA D.20: Período de Ingreso

Vulnerable

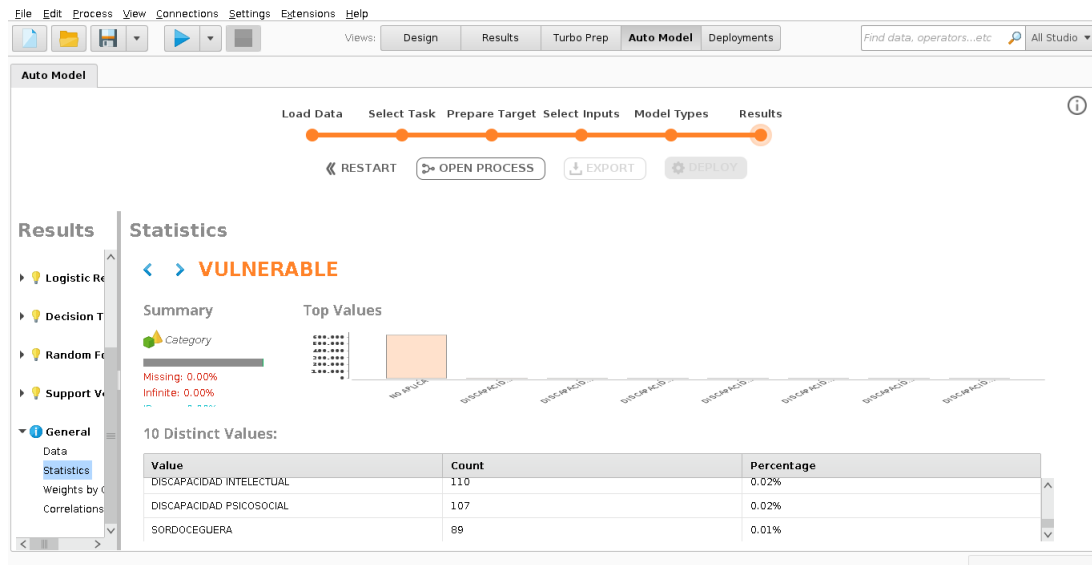


FIGURA D.21: Vulnerable

Año

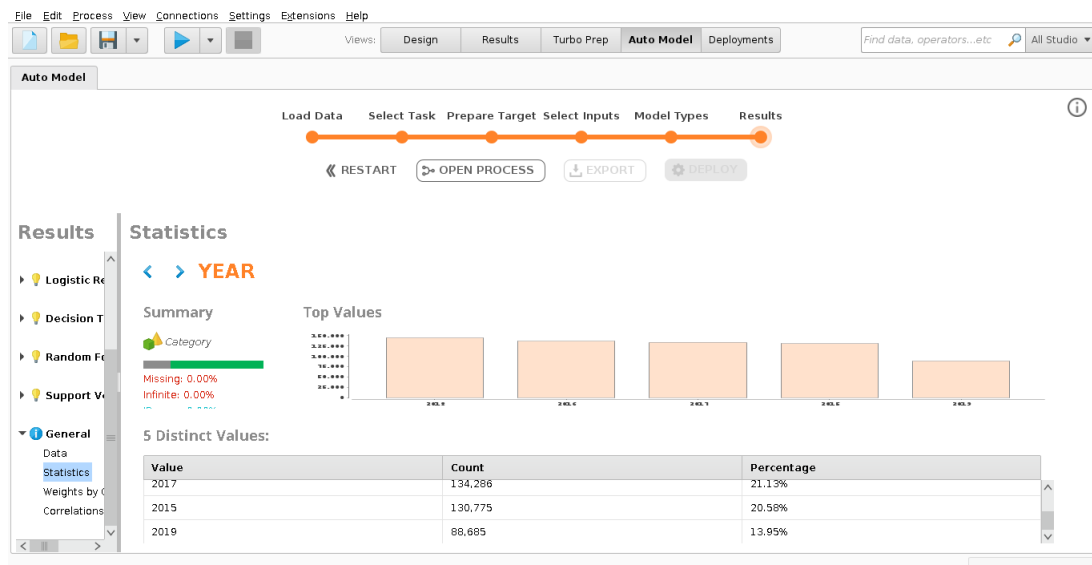


FIGURA D.22: Year

Año de la última matrícula

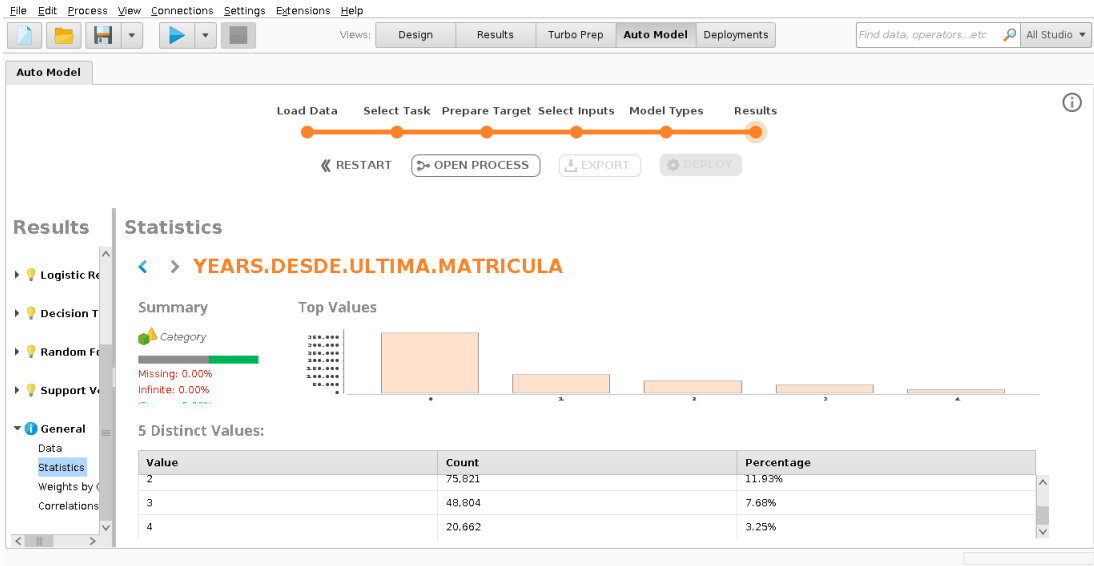


FIGURA D.23: Years.última matrícula

Se tiene un consolidado de los pesos obtenidos para cada atributo en cada método así:

NAIVE BAYES	
ATRIBUTO	Weight
CALIFICACIONES	0,268
ESTADO	0,046
YEARS.DESDE.ULTIMA.MATRICULA	0,026
GENERALIZED LINEAR MODEL	
ATRIBUTO	Weight
ESTRATEGIA	0,254
YEARS.DESDE.ULTIMA.MATRICULA	0,171
CALIFICACIONES	0,138
ESTADO	0,085
ESTRATO	0,072
APOYO	0,061
FACULTAD	0,042
PERIODO	0,035
YEAR	0,027
GENERO	0,025
CONVENIO	0,007
VULNERABLE	0
LOGISTIC REGRESSION	
ATRIBUTO	Weight
CALIFICACIONES	0,203
YEARS.DESDE.ULTIMA.MATRICULA	0,126
ESTADO	0,097
GENERO	0,08
YEAR	0,06
FACULTAD	0,037
ESTRATEGIA	0,012
CONVENIO	0,002
DECISION TREE	
ATRIBUTO	Weight
CALIFICACIONES	0,215
YEARS.DESDE.ULTIMA.MATRICULA	0,022
CONVENIO	0,021
GENERO	0,013
ESTADO	0,003
RANDOM FOREST	
ATRIBUTO	Weight
CALIFICACIONES	0,23
ESTADO	0,137
YEARS.DESDE.ULTIMA.MATRICULA	0,131
SUPPORT VECTOR MACHINE	
ATRIBUTO	Weight
YEARS.DESDE.ULTIMA.MATRICULA for DESERTOR REGISTRADO	0,571
ESTRATO for DESERTOR POR PERIODOS	0,093
ESTADO for DESERTOR POR PERIODOS	0,086

BIBLIOGRAFÍA

- (2018), *Bases del Plan Nacional de Desarrollo 2018-2022: Pacto por Colombia. Pacto por la equidad*, URL <https://colaboracion.dnp.gov.co/CDT/Prensa/PND-2018-2022.pdf>.
- ALFONSO-J., G., A. M. ANTELM-LANZAT, M. L. CACHEIRO-GONZÁLEZ y E. PÉREZ-NAVÍO (2019), «School dropout factors: a teacher and school manager perspective», *Educational Studies*, **45**(6), págs. 756–770, URL <https://doi.org/10.1080/03055698.2018.1516632>.
- APARICIO, E., M. IMELDA, J. KUMUL, E. JOSÉ y E. AVILA (2020), «SOBRE FACTORES INSTITUCIONALES», .
- ARRIETA MATOS, L. (2020), «Remisión estudiantil: estrategia de atención para alertas tempranas desde el aula.», *Congresos CLABES, 1140-1146*, URL <https://revistas.utp.ac.pa/index.php/clabes/article/view/2726>.
- BELL, D. E., R. L. KEENEY y H. RAIFFA (1977), *Conflicting objectives in decisions*, John Wiley & Sons.
- BELTRAN, B. (2016), «MINERÍA DE DATOS (Vol. 30, Issue 1)», .
- BETANCOURT, G. (2005), «LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs)», *Scientia Et Technica*.
- BOSER B. E., I. M. . V. V. N., GUYON (1992), «A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory», , pág. 144–152.

- BOUYSSOU, D. y T. MARCHANT (2015), «On the relations between ELECTRE TRI-B and ELECTRE TRI-C and on a new variant of ELECTRE TRI-B», *European Journal of Operational Research*, **242**(1), págs. 201–211.
- BRUNNER, J. G. C. (2016), *Education in Colombia*, Organización para la Cooperación y el Desarrollo Económicos (OCDE), París.
- CAMARGO GARCÍA, A. J. (2020), *Modelo para la predicción de la deserción de estudiantes de pregrado, basado en técnicas de minería de datos*, Tesis de Maestría, URL <https://hdl.handle.net/11323/7077>.
- CARVAJAL-OLAYA, P. y TREJOS-CARPINTERO (2016), «Revisión De Estudios Sobre Deserción Estudiantil En Educación Superior En Latinoamerica Bajo La Perspectiva De Pierre Bourdieu Universidad Tecnológica de Panamá.», *Congreso CLABES VI, Quito - Ecuador*, URL <https://revistas.utp.ac.pa/index.php/clabes/article/view/1324>.
- CHIN K.S., D. C., LIANG J. ((2003)), *Rough Set Data Analysis Algorithms for Incomplete Information Systems*, URL https://doi.org/10.1007/3-540-39205-X_35.
- CORENA GUTIERREZ, A., P. CONDE y E. MAHECHA (2012), «Definiciones fundamentales sobre deserción estudiantil», *Criterio Jurídico Garantista*, **4**.
- CUNHA, J., E. MOURA y C. ANALIDE (2016), *Data Mining in Academic Databases to Detect Behaviors of Students Related to School Dropout and Disapproval*, págs. 189–198.
- DE LA REPÚBLICA DE COLOMBIA, S. (2016), «LEY 1804 DE 2016», URL <http://www.suin-juriscol.gov.co/viewDocument.asp?ruta=Leyes/30021778>.
- DE PLANEACIÓN UNIVERSIDAD SIMÓN BOLÍVAR, D. (2021), URL https://reportes.unisimon.edu.co/planeacion/Informe_Gestion_Sostenibilidad_Unisimon_2020_final.pdf.

- DELEN, D. (2010), «A comparative analysis of machine learning techniques for student retention management», *Decision Support Systems*, **49**, págs. 498–506.
- DICOVSKIY-RIOBOO, L. M. y M. E. PEDROZA-PACHECO (2018), «Predicción de deserción y éxito en estudiantes. Caso de estudio ingeniería agroindustrial de la UNI Norte, Nicaragua, 2011-2015», *Nexo Revista Científica*, **Vol. 31**(1), págs. 16–27.
- DONOSO, S. y E. SCHIEFELBEIN (2007), «ANÁLISIS DE LOS MODELOS EXPLICATIVOS DE RETENCIÓN DE ESTUDIANTES EN LA UNIVERSIDAD: UNA VISIÓN DESDE LA DESIGUALDAD SOCIAL», *Estudios pedagógicos (Valdivia)*, **33**, págs. 7–27, URL http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-07052007000100001&nrm=iso.
- DRUCKER, P. F. (2004), *"La sociedad postcapitalista"*, primera edición, Norma, Medellín, Colombia.
- EDWARDS, W. y B. FASOLO (2001), «Decision Technology», *Annual review of psychology*, **52**, págs. 581–606.
- FIGUEIRA, J. R., V. MOUSSEAU y B. ROY (2016), «ELECTRE methods», en *Multiple criteria decision analysis*, Springer, págs. 155–185.
- FONTALVO, W. (2014), «Análisis comparativo entre las características más relevantes de deserción estudiantil en el programa de Ingeniería Industrial de la Universidad Autónoma del Caribe. Estudiantes activos en el periodo 2013-01 y desertores académicos de los periodos 2011-01», *Escenarios*, **12**, pág. 96.
- FRENCH, S. (1986), *Decision theory: an introduction to the mathematics of rationality*, Halsted Press.
- FRIEDMAN, D. G. M., NIR. GEIGER (1997), *Bayesian Network Classifiers. Machine Learning*, URL <https://doi.org/10.1023/A:1007465528199>.
- GONZÁLEZ FIEGEHEN, L. (2007), «Repitencia y deserción universitaria en América Latina pg.156 2007-03», .

- GRZYMALA-BUSSE, J. W. (1992), «LERS-a system for learning from examples based on rough sets», en *Intelligent decision support*, Springer, págs. 3–18.
- HAN, K. M. P. J., J. (2012), *Data mining concepts and techniques*.
- HASBUN, T., A. ARAYA y J. VILLALON (2016), «Extracurricular Activities as Dropout Prediction Factors in Higher Education Using Decision Trees», en *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, págs. 242–244.
- HURWITZ, J. y D. KIRSCH (2018), *Machine learning from dummies*, John Wiley Sons, Inc.
- ISSUU (16 de Diciembre de 2014), «https://issuu.com/alfa_guia/docs/proyecto_guia >> , .
- JIA, J.-W. y M. MANOHAR (2014), *Predictive Models for Undergraduate Student Retention Using Machine Learning Algorithms*, tomo 1, págs. 315–329.
- JIMERSON, S., P. FERGUSON y A. WHIPPLE (2002), «Exploring the Association Between Grade Retention and Dropout: A Longitudinal Study Examining Socio-Emotional, Behavioral, and Achievement Characteristics of Retained Students. Contemp School Psychol», , pág. 51–62.
- KALLES, D. y C. PIERRAKEAS (2006), «ANALYZING STUDENT PERFORMANCE IN DISTANCE LEARNING WITH GENETIC ALGORITHMS AND DECISION TREES», *Applied Artificial Intelligence*, **20**(8), págs. 655–674, <https://doi.org/10.1080/08839510600844946>, URL <https://doi.org/10.1080/08839510600844946>.
- KEMPER, L., G. VORHOFF y B. WIGGER (2020), «Predicting student dropout: A machine learning approach», *European Journal of Higher Education*, **10**, págs. 1–20.

- LEE, W., M.-J. LEE y M. BONG (2014), «Testing Interest and Self-Efficacy as Predictors of Academic Self-Regulation and Achievement», *Contemporary Educational Psychology*, **39**.
- MANRIQUE, R., B. PEREIRA NUNES, O. MARINO, M. CASANOVA y T. NURMIKKO-FULLER (2019), «An Analysis of Student Representation, Representative Features and Classification Algorithms to Predict Degree Dropout», págs. 401–410.
- METZ, G. W. (2004), «Challenge and Changes to Tinto's Persistence Theory: A Historical Review», *Journal of College Student Retention: Research, Theory & Practice*, **6**(2), págs. 191–207, <https://doi.org/10.2190/M2CC-R7Y1-WY2Q-UPK5>, URL <https://doi.org/10.2190/M2CC-R7Y1-WY2Q-UPK5>.
- MIGALI, G. y Z. EUGENIO (2017), «Personality traits, forgone health care and high school dropout: Evidence from US adolescents», *Journal of Economic Psychology*, **62**, págs. 98 – 119.
- MINISTERIO DE EDUCACIÓN NACIONAL COLOMBIANO, . (2015), *Guía para la implementación del modelo de gestión para la permanencia y graduación estudiantil en instituciones de educación superior*, Imprenta Nacional de Colombia.
- MONNAPPA, A. (2017), «Data Science vs. Big Data vs. Data Analytics», URL <https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>.
- MORENO-SALINAS, J.-G. (2017), «Científico de datos: codificando el valor oculto e intangible de los datos», *Revista Digital Universitaria RDU UNAM*, **18**, URL <http://www.revista.unam.mx/vol.18/num7/art53/>.
- MOULAY, H., E.-K. HACHEM, C. ZITI y MUSTAPHABASSIRI (2018), «The Use of EDM on Predicting the Rate of Moroccan University Dropouts: Sharia university, Fez as A Case Study», *International Journal of Engineering and Technology*, **Vol 7**(4), págs. 53–56.

- MOUSSEAU, V. y R. SLOWINSKI (1998), «Inferring an ELECTRE TRI model from assignment examples», *Journal of global optimization*, **12**(2), págs. 157–174.
- PAL, S. (2012), «Mining Educational Data to Reduce Dropout Rates of Engineering Students», *International Journal of Information Engineering and Electronic Business*, **4**.
- PATEL, M. R., M. P. VASHI y B. V. BHATT (2017), «SMART-Multi-criteria decision-making technique for use in planning activities», *Proceedings of New Horizons in Civil Engineering (NHCE-2017)*, Surat India.
- PAWLAK, Z. (1998), «Reasoning about Data — A Rough Set Perspective», .
- PAWLAK, Z. (2012), *Rough sets: Theoretical aspects of reasoning about data*, tomo 9, Springer Science & Business Media.
- PIMPA, C. (2013), «Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program», en *International MultiConference of Engineers and Computer Scientists*, tomo 1.
- QUINLAN, J. R. (1986), *Induction of decision trees. Machine Learning*, tomo 1.
- RAJ, V. y M. VANNAN (2020), «Predicting Student Failure in University Examination using Machine Learning Algorithms», , págs. 2278–3075.
- RAUSCH, J. L. y M. HAMILTON (2006), «Goals and Distractions: Explanations of Early Attrition from Traditional University Freshmen.», .
- ROBINSON, S., J. JAGGERS, J. RHODES, B. BLACKMON y W. CHURCH II (2016), «Correlates of educational success: Predictors of school dropout and graduation for urban students in the Deep South», *Children and Youth Services Review*, **73**.
- ROLLINS, J. (2015), «Metodología Fundamental para la Ciencia de Datos», URL <https://www.ibm.com/downloads/cas/6RZMKDN8>.
- ROWTO, V. (2017), «Early Detection of At-Risk Undergraduate Students through Academic Performance Predictors», *Higher Education Studies*, **7**, pág. 42.

- ROY, B. (), «The outranking approach and the foundations of ELECTRE methods», en *Readings in multiple criteria decision aid*, Springer, págs. 155–183.
- SAATY, T. L. (2004), «Decision making—the analytic hierarchy and network processes (AHP/ANP)», *Journal of systems science and systems engineering*, **13**(1), págs. 1–35.
- SIMON, H. (1960), *The New Science of Management Decision*, Ford distinguished lectures, Harper, URL <https://books.google.com.co/books?id=nktqAAAAMAAJ>.
- SOLIS, M., T. MOREIRA-MORA, R. LOAIZA, T. FERNÁNDEZ-MARTÍN y M. HERNÁNDEZ-JIMÉNEZ (2018), «Perspectives to Predict Dropout in University Students with Machine Learning», .
- SPADIES (2020), «Estadísticas de Deserción», URL <https://www.mineducacion.gov.co/sistemasinfo/spadies/Informacion-Institucional/357549:Estadisticas-de-Desercion>.
- STRAUSS, B., W. STRAUSS y N. HOWE (1991), *Generations: The History of America's Future, 1584 to 2069*, William Morrow Company, New York.
- TINTO, V. (1975), «Dropout from Higher Education: A Theoretical Synthesis of Recent Research», *Review of Educational Research*, **45**(1), págs. 89–125, URL <https://doi.org/10.3102/00346543045001089>.
- TINTO, V. y B. PUSSEY (2006), «Moving From Theory to Action: Building a Model of Institutional Action for Student Success», .
- TUKEY, J. W. (1962), «The Future of Data Analysis», *Ann. Math. Statist.*, **33**(1), págs. 1–67, URL <https://doi.org/10.1214/aoms/1177704711>.
- VANDER-AALST, W. (2014), *Data Scientist: The Engineer of the Future*, tomo "7", "Springer, Cham", págs. 13–26.
- VICERRECTORÍA, U. S. B. (2018), «Modelo de permanencia y graduación estudiantil USB», .

- VINCKE, P. (1992), *Multicriteria Decision-Aid*.
- WATSON, G., G. G. JOHNSON@GRIFFITH.EDU.AU y H. AUSTIN (2004), «Exploring relatedness to field of study as an indicator of student retention», *Higher Education Research Development*, **23**.
- YEVSEYEVA, I. (2007), *SOLVING CLASSIFICATION PROBLEMS WITH MULTI-CRITERIA DECISION AIDING APPROACHES*.
- ZOPOUNIDIS, C. y M. DOUMPOS (2002), «Multicriteria classification and sorting methods: A literature review», *European Journal of Operational Research*, **138**(2), págs. 229–246.